INTRODUCTION TO COMPEX NETWORKS

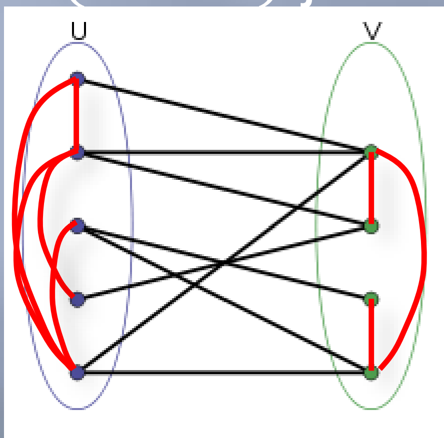
János Kertész

Janos.Kertesz@gmail.com

4. ERDŐS-RÉNYI AND WATTS-STROGATZ GRAPHS

Addendum: Projection of bipartite (two mode) networks

$G_{\text{bipartite}}(U, V, E)$, where $|U| = N$ and $|V| = M$. As there are no links between nodes within *U* (or within *V*) the $(N + M) \times (N + M)$ joint adjacency matrix $\boldsymbol{A}(U, V)$ will be:



N=5 *M*=4

$$\begin{array}{c} \\ U \\ \\ V \end{array} \begin{array}{c} \overbrace{\qquad\qquad\qquad}^{U} \overbrace{\qquad\qquad}^{V} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Enough: the *NXM* (or *MXN*) matrix

Projection onto *U*:

$$\boldsymbol{A}(U) = \boldsymbol{A}(U, V)\boldsymbol{A}^T(U, V) = \boldsymbol{A}(U, V)\boldsymbol{A}(V, U)$$

Onto *V*:

$$\boldsymbol{A}(V) = \boldsymbol{A}(V, U)\boldsymbol{A}^T(V, U) = \boldsymbol{A}(V, U)\boldsymbol{A}(U, V)$$

$$A(V)_{ij} = \sum_k A(U, V)_{ik} A(V, U)_{kj}$$

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 0 & 1 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 2 \end{pmatrix}$$

Projection to *V*

With multiplicities!

A_{ii} = # links to *U* at *i*

A_{ij} = # connections between *i* and *j* $(i \neq j)$

Modeling networks

As technology advances: we a) get access to b) create large networks

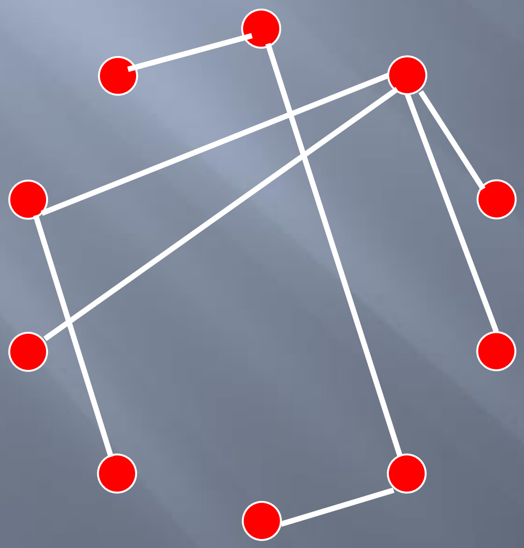
How to model the observed networks?

We can easily generate regular networks (e.g., lattices) but in real networks there is usually a large amount of randomness.

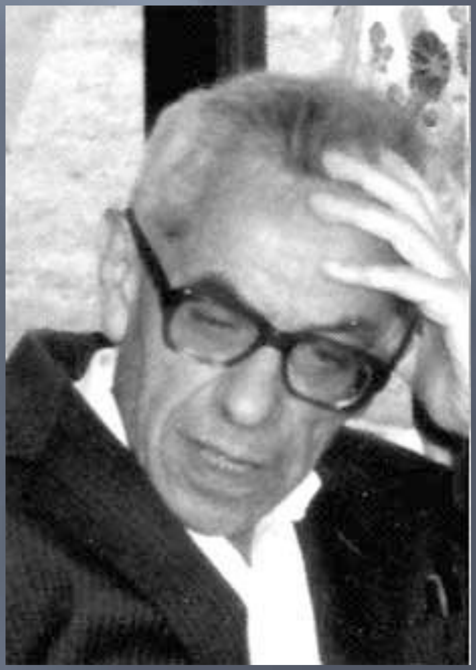
Take the opposite view: Generate the most random network!

Erdös-Rényi model

If we do not know anything else than the number *N* of nodes and the number *L* of links, the simplest thing to do is to put the links at random (no correlations)



N =10, *L*= 8



Paul (Pál) Erdős

Alfréd Rényi

Erdös-Rényi model

This is one realization of the (*N*,*L*) E-R model. The links can be put in many different ways. This defines an ensemble of graphs: the E-R model.

Probabilistic definition:

possible links: $L_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$

probability of having a link between any two nodes:

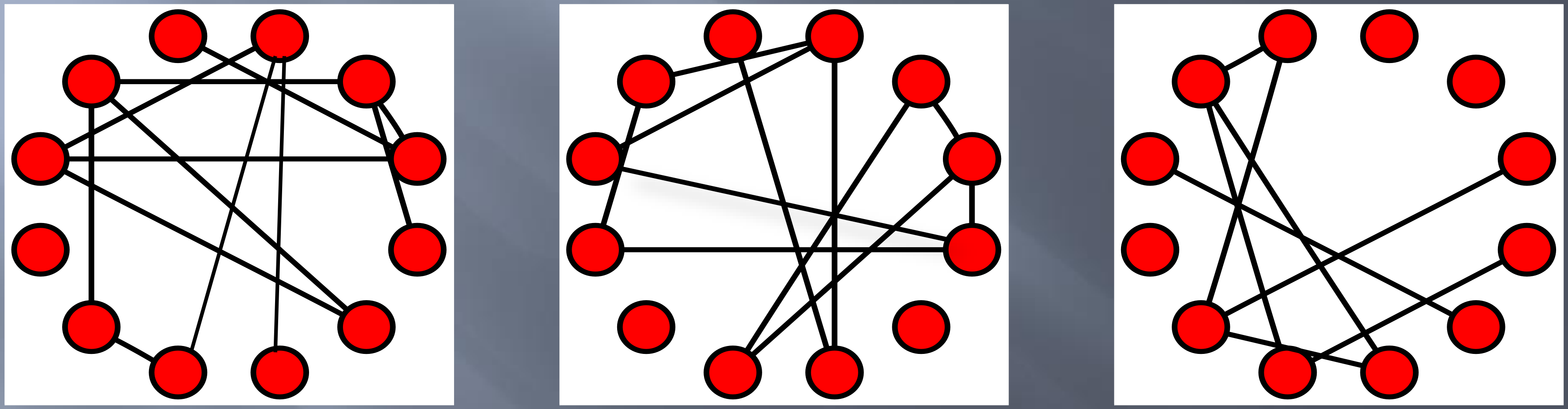
$$p = \langle L \rangle / \frac{N(N-1)}{2} = \frac{2\langle L \rangle}{N(N-1)}$$

This is the $G(N, p)$ model

Erdös-Rényi model

N =12
p =1/6

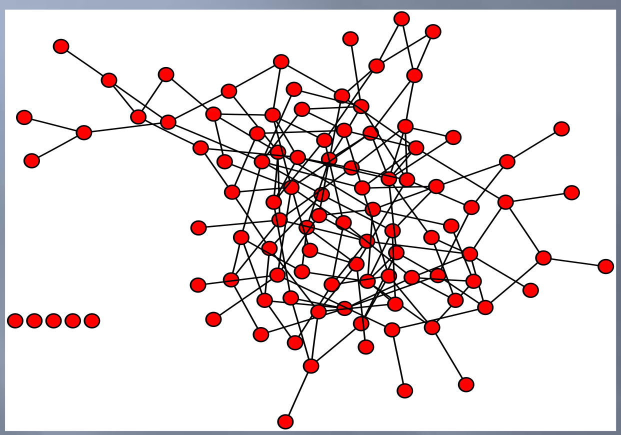
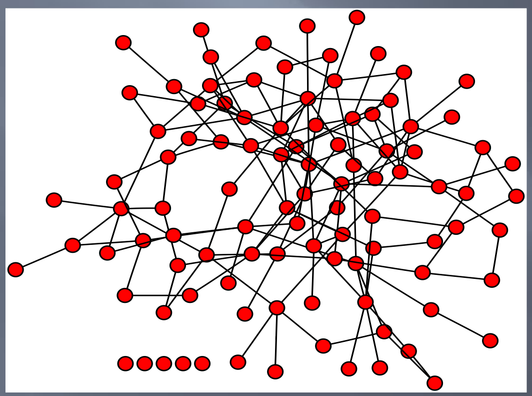
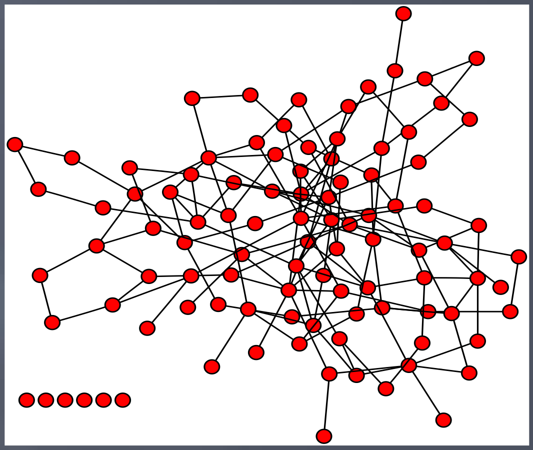
L_{max} = 11*6=66 $\langle L \rangle$ = 11



L = 12 L = 11 L = 8

Erdös-Rényi model

$N = 100$
$p = 0.03$

The probability of one particular configuration *G*

$$P\big(G(N,p;L)\big) = p^L (1-p)^{L_{\max}-L}$$

due to independence

Erdös-Rényi model

How many ways can we put L links on $L_{\max}$ places? (Combinatorics)

If we have a sequence of *L* links, then the first can be put on $L_{\max}$ places, the second on $L_{\max}$ –1, the third on $L_{\max}$ – 2, … the last on, leading to

$$L_{\max}(L_{\max} - 1)(L_{\max} - 2)\cdots(L_{\max} - L + 1)$$

different possibilities. However, the sequence does not matter, thus we have to divide the result by the number of different sequences, which is

$$1 \times 2 \times 3 \times \cdots \times L = L!$$

Erdös-Rényi model

How many ways can we put L links on $L_{\max}$ places?

The result is:

$$\frac{L_{\max}(L_{\max}-1)(L_{\max}-2)\cdots(L_{\max}-L+1)}{L!} = \frac{L_{\max}!}{L!(L_{\max}-L)!} = \binom{L_{\max}}{L}$$

The probability of finding a graph with exactly L links:

$$P(N,p;L) = \binom{L_{\max}}{L} p^L (1-p)^{L_{\max}-L}$$

Binomial distribution

Binomials

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

$P(N,p;L)$ is normalized

Erdös-Rényi model

We show that

$$\langle L \rangle = pL_{\max} \quad \text{and} \quad \langle L^2 \rangle = pL_{\max} + p^2 L_{\max}(L_{\max} - 1)$$

$$\langle L \rangle = \sum_L L \binom{L_{\max}}{L} p^L (1-p)^{L_{\max}-L} =$$

$$\sum_L L \binom{L_{\max}}{L} p^L q^{L_{\max}-L} = p\frac{d}{dp}\sum_L \binom{L_{\max}}{L} p^L q^{L_{\max}-L} =$$

$$p\frac{d}{dp}(p+q)^{L_{\max}} = pL_{\max}$$

Erdös-Rényi model

$$\langle L^2 \rangle = \sum_L L^2 \binom{L_{\max}}{L} p^L (1-p)^{L_{\max}-L} =$$

$$\sum_L L^2 \binom{L_{\max}}{L} p^L q^{L_{\max}-L} = p\frac{d}{dp} p\frac{d}{dp} \sum_L \binom{L_{\max}}{L} p^L q^{L_{\max}-L} =$$

$$p\frac{d}{dp} p\frac{d}{dp} (p+q)^{L_{\max}} = p\frac{d}{dp} pL_{\max}(p+q)^{L_{\max}-1} =$$

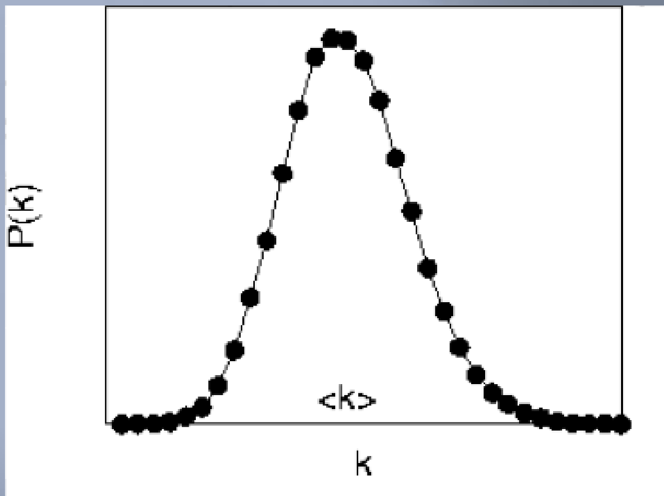
$$pL_{\max} + p^2 L_{\max}(L_{\max}-1)$$

$$s^2 = \langle L^2 \rangle - \langle L \rangle^2 = pL_{\max} + p^2 L_{\max}(L_{\max} - 1) - p^2 L_{\max}^2 =$$

$$L_{\max} p(1 - p)$$

Erdös-Rényi model

Degree distribution



$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

Select k nodes from N-1

probability of having *k* edges

probability of missing N-1-k edges

$$< k >= p(N - 1)$$

$$s_k^2 = p(1 - p)(N - 1)$$

$$\frac{\sigma_k}{< k >} = \left[\frac{1-p}{p} \frac{1}{(N-1)} \right]^{1/2} \approx \frac{1}{(N-1)^{1/2}}$$

As the network size increases, the distribution becomes more and more narrow — the degree of a node is with high probability in the vicinity of <k>.

Probability generating function

Given a distribution $P(i)$ with $i = 0,1,2$ ... the generating function is defined as $G(x) = \sum_i P(i)x^i$.

$P(i) = \frac{1}{i!}\frac{d^iG(x)}{dx^i}\Big|_{x=0}$ thus $G(x)$ is equivalent to $P(i)$.

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} \quad \rightarrow \quad \frac{e^{-\lambda}\lambda^i}{k!}$$

$\lambda = p(N-1) = \langle k \rangle$

$$G_{\text{bin}}(x;p,N) \equiv \sum_{k=0}^{N} \left[\binom{N}{k} p^k (1-p)^{N-k}\right] x^k = \left[1 + (x-1)p\right]^N$$

$$\lim_{\substack{N\to\infty \\ p\to 0}} G_{\text{bin}}(x;p,N) = \lim_{N\to\infty} \left[1 + \frac{\lambda(x-1)}{N}\right]^N = e^{\lambda(x-1)} = \sum_{k=0}^{\infty} \left[\frac{e^{-\lambda}\lambda^k}{k!}\right] x^k$$

$$= G_{\text{Poisson}}(x;\lambda)$$

Erdös-Rényi model

Approximation to the binomial distribution for large *N* and fixed <k> (meaning small *p*).

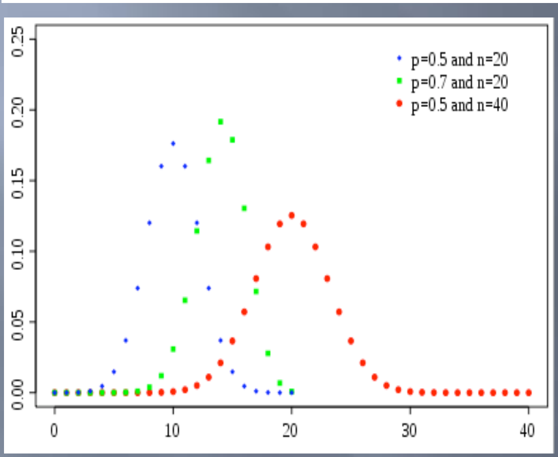
Exact Result

-binomial distribution-

Large N limit

-Poisson distribution-

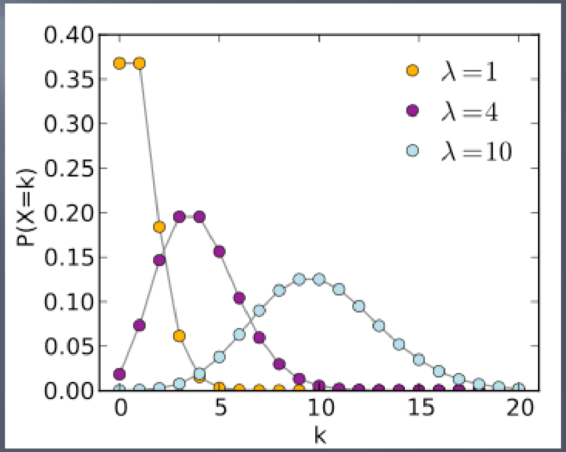
$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} \quad \rightarrow \quad \frac{e^{-<k>} <k>^i}{k!}$$



$$<k> = (N-1)p$$

$$<k^2> = p(1-p)(N-1) + p^2(N-1)^2$$

$$s_k = (<k^2> - <k>^2)^{1/2} = [p(1-p)(N-1)]^{1/2}$$



$$<k> = <k>$$

$$<k^2> = <k>(1 + <k>)$$

$$s_k = (<k^2> - <k>^2)^{1/2} = <k>^{1/2}$$

Erdös-Rényi model

Poisson distribution is narrow:

The chance to have degree much larger than $\langle k \rangle$ is VERY small:

For $k=10$ it is $P(k>10\langle k\rangle) < 2X10^{-13}$

What about reality?

Erdös-Rényi model

According to sociological research, for a typical individual k ~1,000

The probability to find an individual with degree k>2,000 is 10^{-27}.

Given that N ~10^9, the chance of finding an individual with 2,000 acquaintances is so tiny that such nodes are virtually non-existent in a random society.

→a random society would consist of mainly average individuals, with everyone with roughly the same number of friends.

→It would lack outliers, individuals that are either highly popular or recluse.

Erdös-Rényi model

ER fails!

Why do we study it?

Basic reference model: The absolute random limit.

Many properties can be calculated – good playground to test tools.

Erdős-Rényi model

What do we know?

$$P(N,p;L)=\binom{L_{\max}}{L}p^{L}(1-p)^{L_{\max}-L}$$

Binomial distribution

$$\langle L\rangle = pL_{\max} \quad \text{and} \quad \langle L^2\rangle = pL_{\max} + p^2L_{\max}(L_{\max} - 1)$$

$$P(k)=\binom{N-1}{k}p^{k}(1-p)^{(N-1)-k}$$

$$<k>= p(N - 1)$$

$$s_k^2 = p(1- p)(N - 1)$$

$$\frac{\sigma_k}{<k>}=\left[\frac{1-p}{p}\frac{1}{(N-1)}\right]^{1/2}\approx\frac{1}{(N-1)^{1/2}}$$

Sharp distribution

Erdös-Rényi model

Clustering coefficient *C*

The prob. that two neighbors are neighbors of each other. Since in the ER model the probability of a link is always *p*, we have

$$C = p = \langle L \rangle / \frac{N(N-1)}{2} = \frac{2\langle L \rangle}{N(N-1)} = \frac{\langle k \rangle}{N-1}$$

Small!

In (large) social networks there are plenty of triangles!

(Another problem with ER!)

Erdős-Rényi model

Percolation transition (in a complete graph)

If p is very small – only small isolates

If p is large – giant component (+ isolates)

Where is the transition?

u prob. that a randomly chosen node does NOT belong to the giant component. (Either not connected or connected to a node not connected to the giant)

$$u = (1 - p + pu)^{N-1} = \left[1 - \frac{\langle k \rangle}{N-1}(1-u)\right]^{N-1}$$

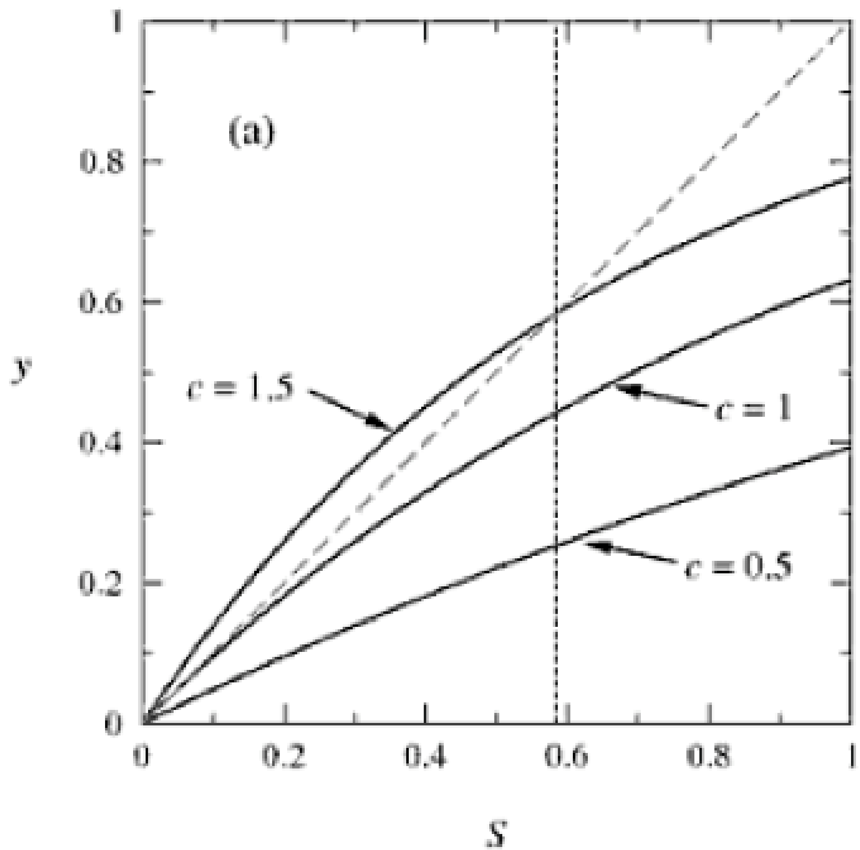
Tree property! Giant comp. near crit. point

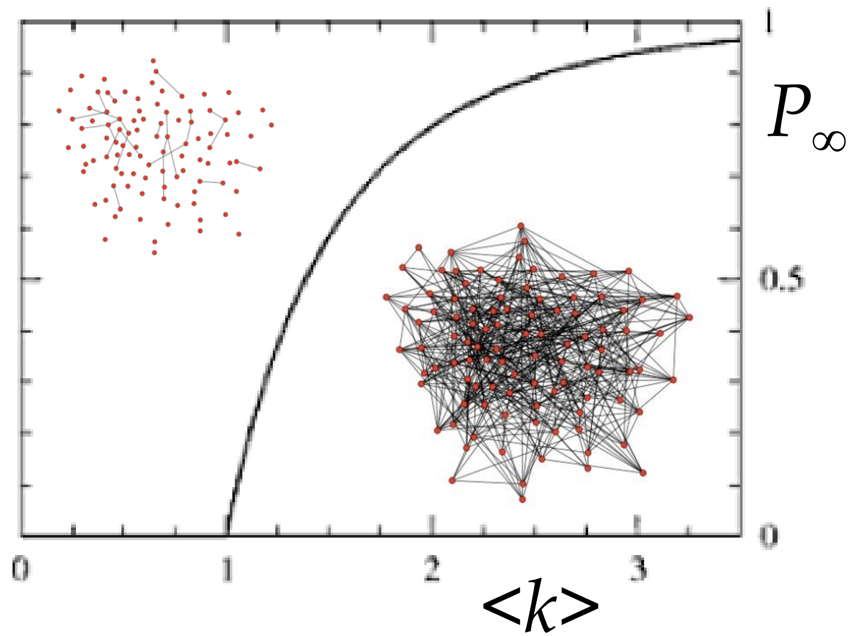
$$\ln u = (N-1)\ln\left[1 - \frac{\langle k \rangle}{N-1}(1-u)\right] \approx -\langle k \rangle (1-u) \qquad u = e^{-\langle k \rangle (1-u)}$$

$$P_\infty = 1 - u = 1 - e^{-\langle k\rangle P_\infty}$$

Is the prob. that a node belongs to the giant comp.

Graphical solution:





$$\frac{d}{dP_\infty}\left(1 - e^{-\langle k\rangle P_\infty}\right) = 1 \quad \Rightarrow \quad \langle k\rangle e^{-\langle k\rangle P_\infty} = 1$$

At transition $P_\infty = 0$

$\langle k\rangle_c = 1$

Newmann book

At transition $P_\infty = 0$ and $<k>_c = 1$

How does P_∞ depend on $\varepsilon = <k> - <k>_c = <k> - 1$?

$$P_\infty = 1 - u = 1 - e^{-<k>P_\infty}$$

Near the transition P_∞ is still small. Let $y = \langle k \rangle P_\infty$

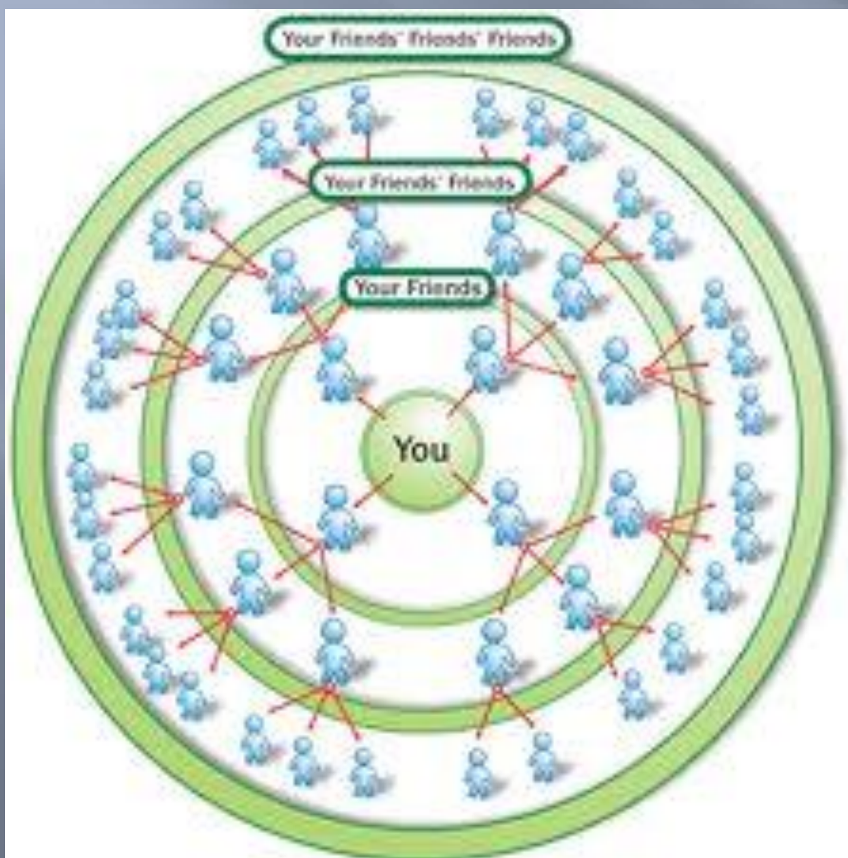
$$P_\infty = 1 - e^{-<k>P_\infty} \Rightarrow y = \langle k \rangle \left(1 - e^{-y}\right)$$

$$y = \langle k \rangle \left(y - \frac{1}{2} y^2 \right)$$

$$\langle k \rangle - 1 = \langle k \rangle \frac{1}{2} y \Rightarrow P_\infty \sim \varepsilon^\beta \text{ with } \beta = 1$$

Erdős-Rényi model

$P(k)$ is narrow and for small $< k > -1$ there are no loops. Let us substitute the network nodes such that they all have the degree $<k>$ (even if this is not an integer). The average # nodes at (geodesic) distance *s* is $<k>^s$.



Exponential network:

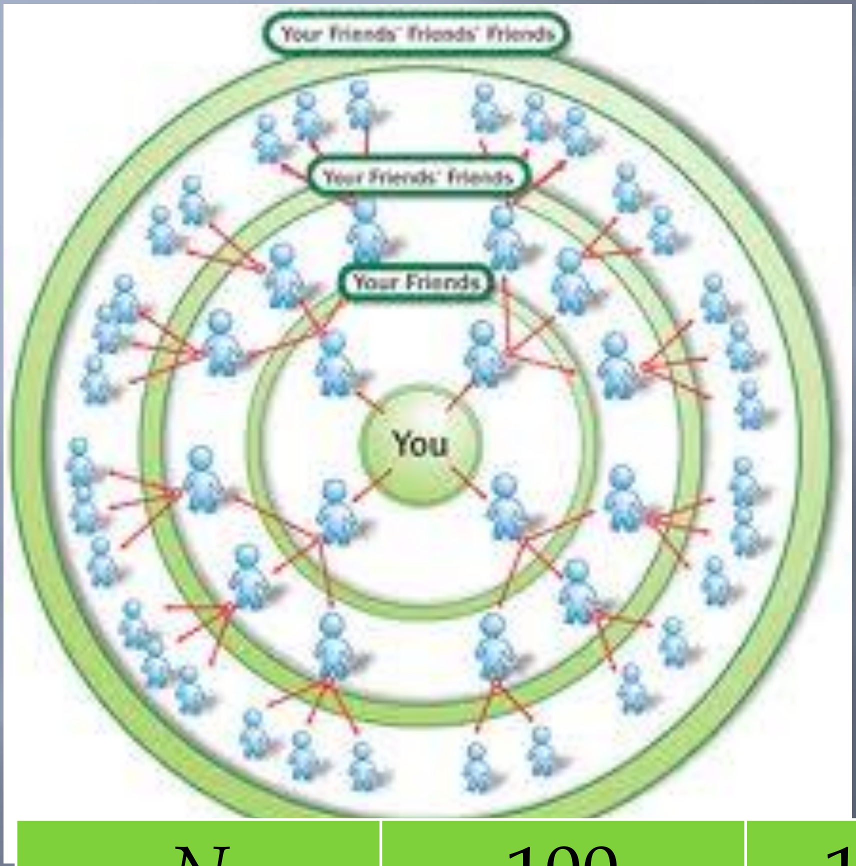
$$N = \frac{\langle k \rangle^{s+1} - 1}{\langle k \rangle - 1} = \begin{cases} \to \infty \text{ as } s \to \infty \text{ if } \langle k \rangle > 1 \\ \to \dfrac{1}{1 - <k>} \text{ as } s \to \infty \text{ if } \langle k \rangle < 1 \end{cases}$$

For $<k> < 1$ no giant component. $N \approx <k>^s$

For $<k> > 1$ there is giant component.

Erdős-Rényi model

Pathlengths proportional to s



$$\langle k \rangle^s » N \quad \text{or} \quad s » \frac{\ln N}{\ln \langle k \rangle}$$

The length of the average distance grows only logarithmically with N

N	100	10,000	1,000,000	10^a
s	2	4	6	a

Small world

Erdős-Rényi model

ER model (large *n* limit)

- Random graph (ensemble)
- No correlations
- Sharp distribution of degrees
- Small clustering coefficient
- Percolation transition at $\langle k \rangle = 1$
- Small world properties (in the giant component)

This is the model. What about reality?

Erdős-Rényi model

As quantitative data about real networks become available, we can compare their topology with the predictions of random graph theory. Note that once we have N and <k> for an ER random network, from it we can derive every measurable property. Indeed, we have:

Average distance:

$$< d_{rand} > \approx \frac{\log N}{\log \langle k \rangle}$$

Clustering Coefficient:

$$C_{rand} = p = \frac{\langle k \rangle}{N}$$

Degree Distribution:

$$P_{rand}(k) \cong C_{N-1}^{k} p^{k} (1 - p)^{N-1-k}$$

$$P(k) = e^{-<k>} \frac{< k >^{k}}{k!}$$

Erdős-Rényi model

Average distance:

Prediction:

$$< d_{rand} > \approx \frac{\log N}{\log\langle k\rangle}$$

Real networks have short distances like ER random graphs.

Data:

- Food web
- Neural network
- Collaboration networks
- WWW
- Metabolic networks
- Internet

All small worlds!

Erdős-Rényi model

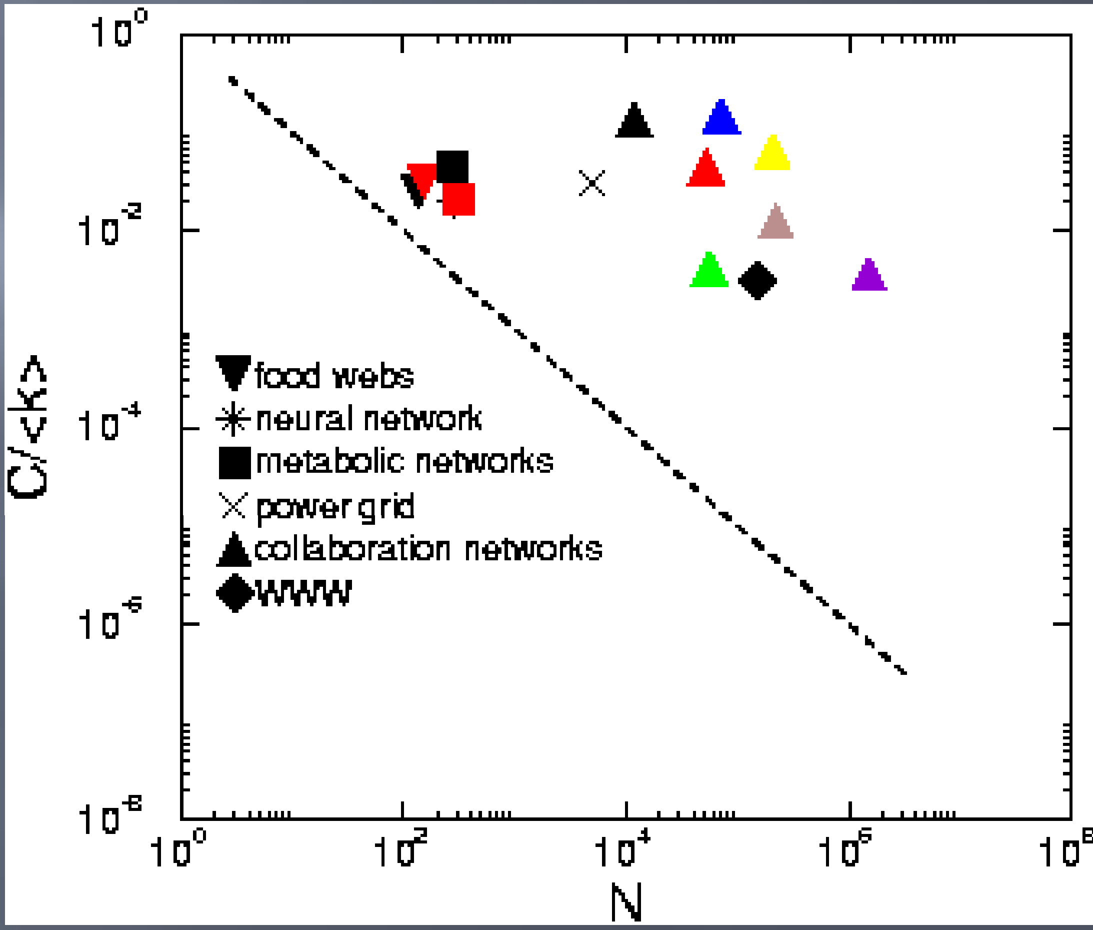
Clustering coefficient

Prediction:

$$C_{rand} = \frac{\langle k \rangle}{N}$$

C_{rand} underestimates the clustering coefficient of real networks
by orders of magnitudes!

Real world NW-s



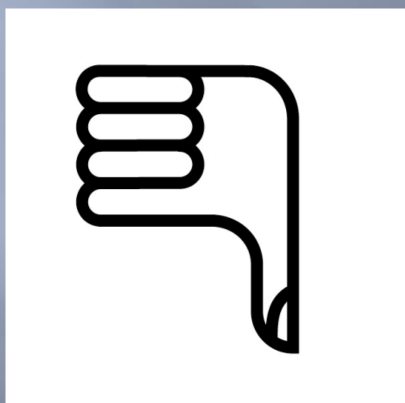
Barabasi

Erdős-Rényi model

Degree distribution

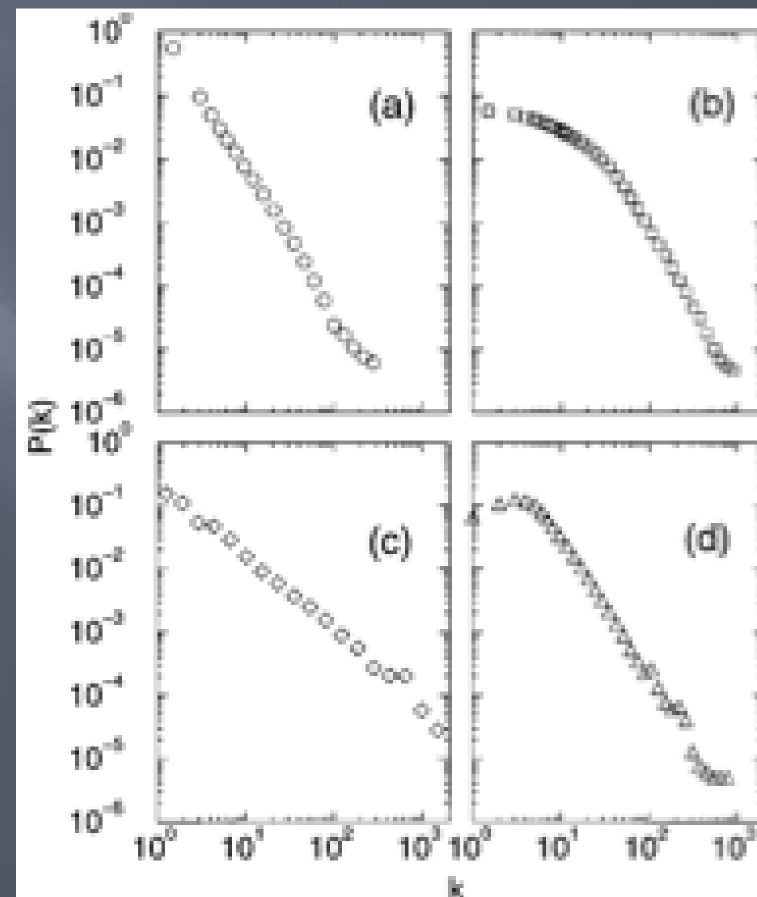
Prediction:

$$P_{rand}(k) @ C_{N-1}^{k} p^{k} (1- p)^{N-1-k}$$



Data:

(a)Internet;
(b) Movie Actors;
(c)Coauthorship, high energy physics;
(d) Coauthorship, neuroscience

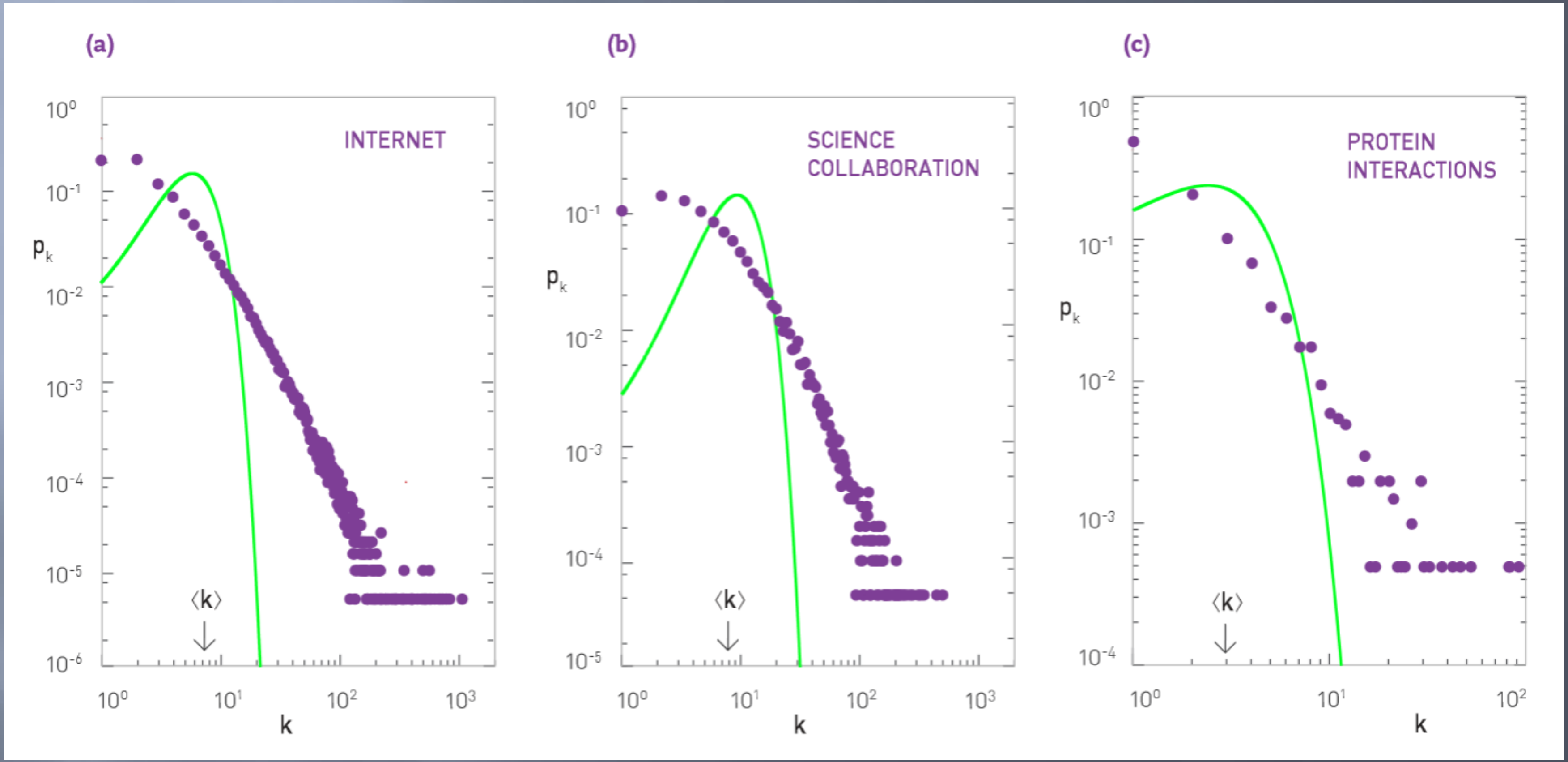


Data follow power laws

$$P(k) » k^{-9}$$

Barabasi

Degree distribution



Green line: Poisson

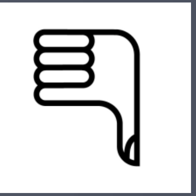
Erdős-Rényi model

By comparing the measures as obtained from the data and from the model, it becomes clear that the model is far from reality. No real system is properly described by it.

Average distance: $< d_{rand} > \approx \dfrac{\log N}{\log\langle k\rangle}$

Clustering Coefficient: $C_{rand} = p = \dfrac{\langle k\rangle}{N}$

Degree Distribution: $P_{rand}(k) \cong C_{N-1}^{k} p^{k} (1 - p)^{N-1-k}$



It seems to capture the small world property!

Barabasi

Some further features

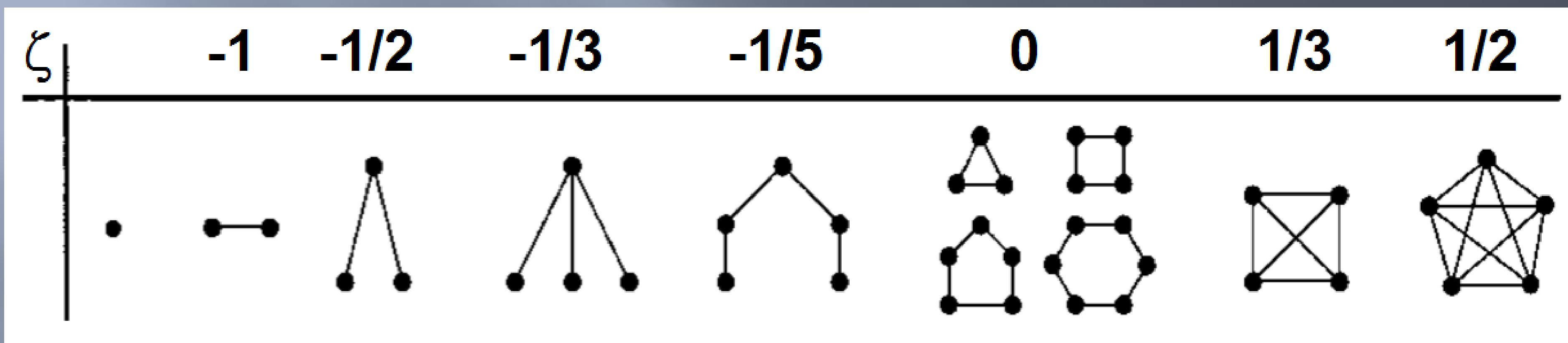
$$|\text{Largest component}| \begin{cases} \sim \log N \text{ if } \langle k \rangle < 1 & \text{subcritical} \\ \sim N^{2/3} \text{ if } \langle k \rangle = 1 & \text{critical} \\ \sim N \text{ if } \langle k \rangle > 1 & \text{supercritical} \end{cases}$$

Probability that a randomly selected node belongs to the largest component → 0 for $\langle k \rangle \leq 1$ when $N \to \infty$.

There is one and only one largest component (giant component) $\langle k \rangle \geq 1$ when $N \to \infty$.

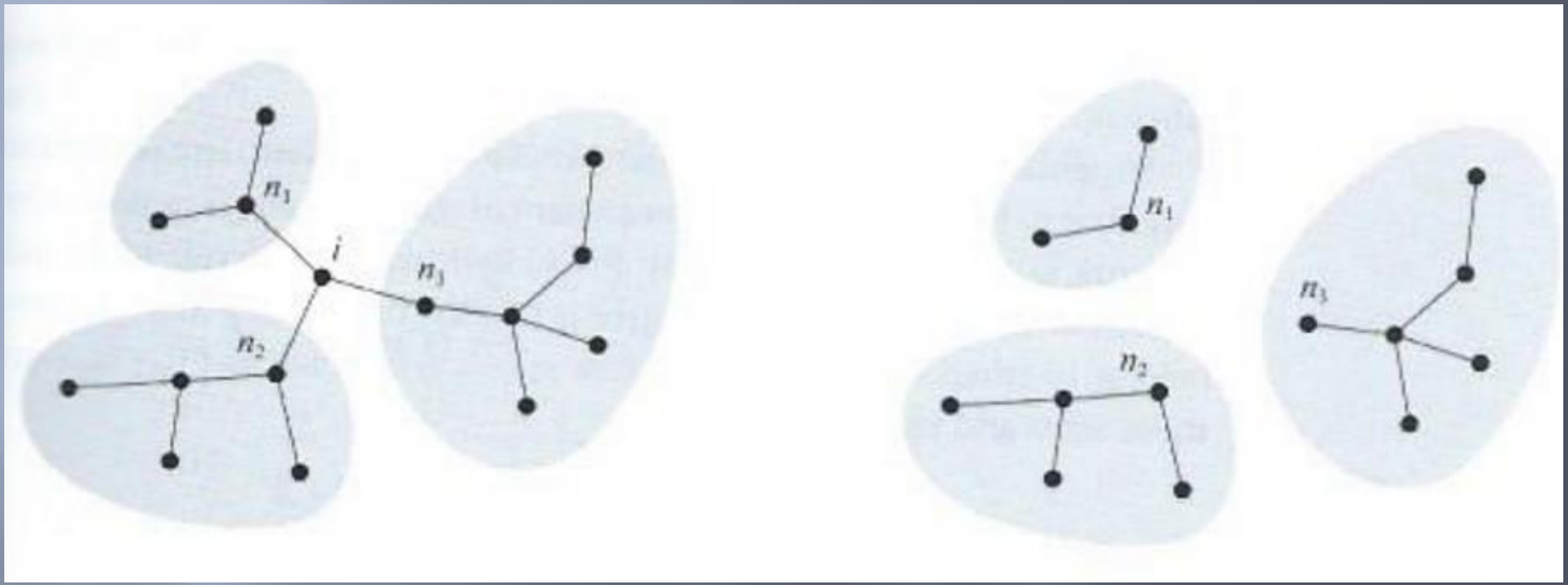
Some further features

It is possible to calculate the critical N-dependence of $\langle k \rangle$ below which $\langle|\text{subgraph}|\rangle \rightarrow 0$ when $N \rightarrow \infty$.



If $\mathcal{O}(\langle k \rangle) < N^{\zeta}$ then $\langle|\text{subgraph}|\rangle \rightarrow 0$;
if $\mathcal{O}(\langle k \rangle) > N^{\zeta}$ then $\langle|\text{subgraph}|\rangle > 0$ when $N \rightarrow \infty$.

Finite components



Prob. that a node belongs to component of size s: p_s

Finite components are trees:

$$P(s|k) = \sum_{s_1} \dots \sum_{s_k} \prod_{j=1}^{k} p_{s_j} \delta\left(s-1, \sum_j s_j\right)$$

Finite components

$$P(s|k) = \sum_{s_1} \dots \sum_{s_k} \prod_{j=1}^{k} p_{s_j} \delta\left(s-1, \sum_j s_j\right)$$

$$p_s = \sum_{k=1}^{\infty} P(k)\, P(s|k) = \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \sum_{s_1} \dots \sum_{s_k} \prod_{j=1}^{k} p_{s_j} \delta\left(s-1, \sum_j s_j\right)$$

Generating function $g(z) = \sum_{s=1}^{\infty} p_s z^s$

$$g(z) = \sum_{s=1}^{\infty} z^s e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \sum_{s_1} \dots \sum_{s_k} \prod_{j=1}^{k} p_{s_j} \delta\left(s-1, \sum_j s_j\right)$$

Finite components

$$g(z) = \sum_{s=1}^{\infty} z^s e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \sum_{s_1} \dots \sum_{s_k} \prod_{j=1}^{k} p_{s_j} \delta\left(s-1, \sum_j s_j\right)$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \sum_{s_1} \dots \sum_{s_k} \prod_{j=1}^{k} p_{s_j} z^{1+\sum_j s_j} =$$

$$e^{-\lambda} z \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \sum_{s_1} \dots \sum_{s_k} \prod_{j=1}^{k} \left[p_{s_j} z^{s_j}\right] =$$

$$e^{-\lambda} z \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \left(\sum_{s} p_s z^s\right)^k = z\, exp(\lambda(g(z)) - 1) = g(z)$$

Some further features

For $\langle \mathrm{k} \rangle > 1$, the average size of small (not giant) components gets independent of N.
The average size χ of finite clusters is:

$$\chi = \frac{\langle s p_s \rangle}{\langle p_s \rangle} = \frac{1}{1 - \langle k \rangle + \langle k \rangle P_\infty}$$

(for derivation see Newman's book)

where p_s is the prob. that a randomly chosen node belongs to a finite component of size *s*, and P_∞ is the relative weight of the giant component.

As P_∞ vanishes as $\varepsilon = 1 - \langle k \rangle$, we have $\chi \sim \varepsilon^{-\gamma}$ with $\gamma = 1$.

Summary

Erdős-Rényi model:

- Simplest
- No correlations
- Many features calculable
- Sharp degree distribution (Poisson for large N)
- $C \to 0$ as $N \to \infty$.
- Percolation transition at $\langle k \rangle = 1, \beta = 1, \gamma = 1$.
- Small world for $\langle k \rangle > 1$ due to exponential graph structure.
- Basic reference model

Small world

A fascinating game grew out of this discussion. One of us suggested performing the following experiment to prove that the population of the Earth is closer together now than they have ever been before. We should select any person from the 1.5 billion inhabitants of the Earth – anyone, anywhere at all. He bet us that, using no more than *five* individuals, one of whom is a personal acquaintance, he could contact the selected individual using nothing except the network of personal acquaintances. For example, "Look, you know Mr. X.Y., please ask him to contact his friend Mr. Q.Z., whom he knows, and so forth."

I proposed a more difficult problem: to find a chain of contacts linking myself with an anonymous riveter at the Ford Motor Company — and I accomplished it in four steps. The worker knows his foreman, who knows Mr. Ford himself, who, in turn, is on good terms with the director general of the Hearst publishing empire. I had a close friend, Mr. Árpád Pásztor, who had recently struck up an acquaintance with the director of Hearst publishing. It would take but one word to my friend to send a cable to the general director of Hearst asking him to contact Ford who could in turn contact the foreman, who could then contact the riveter, who could then assemble a new automobile for me, should I need one.

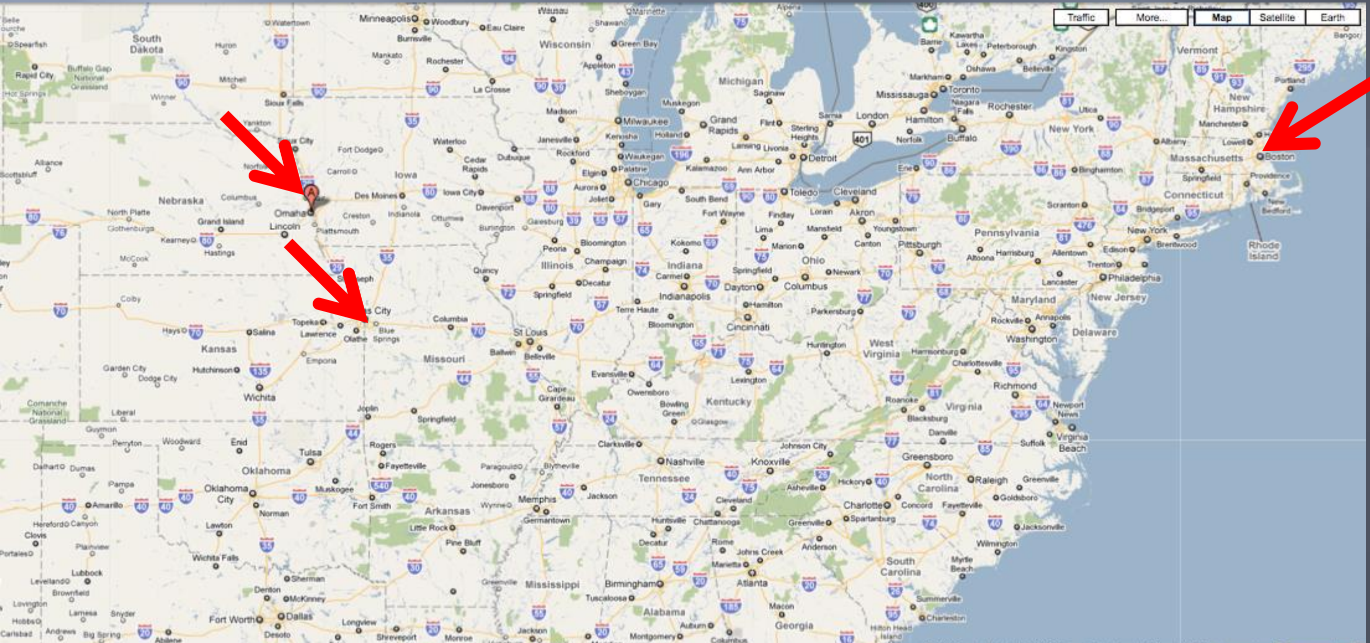
Frigyes Karinthy: Chains (1929)

Small world

More scientifically:

Stanley Milgram experiment

He gave letters addressed to a Boston broker to people in the Midwest and asked them to hand them to acquaintees such that using only personal contacts the letters should find the broker as soon as possible.



Small world

The rules of the Milgram game:

1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.

2. DETACH ONE POSTCARD. FILL IT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

3. IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.

4. IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POST CARDS AND ALL) TO A PERSONAL ACQUAINTANCE WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may send the folder to a friend, relative or acquaintance, but it must be someone you know on a first name basis.


New Hampshire
Vermont
Massachusetts
Maine
North Dakota
Minnesota
Wisconsin
South Dakota
Michigan
New York
Rhode Island
Connecticu
Pennsylvania
Nebraska
Iowa
Illinois
Indiana
Ohio
New Jersey
start
Delaware
Kansas
Missouri
Kentucky
Virginia
Maryland
West Virginia
North Carolina
Tennessee
Oklahoma
Arkansas
South Carolina
Alabama
Georgia
Mississippi
Texas
Louisiana
Florida

Wikipedia

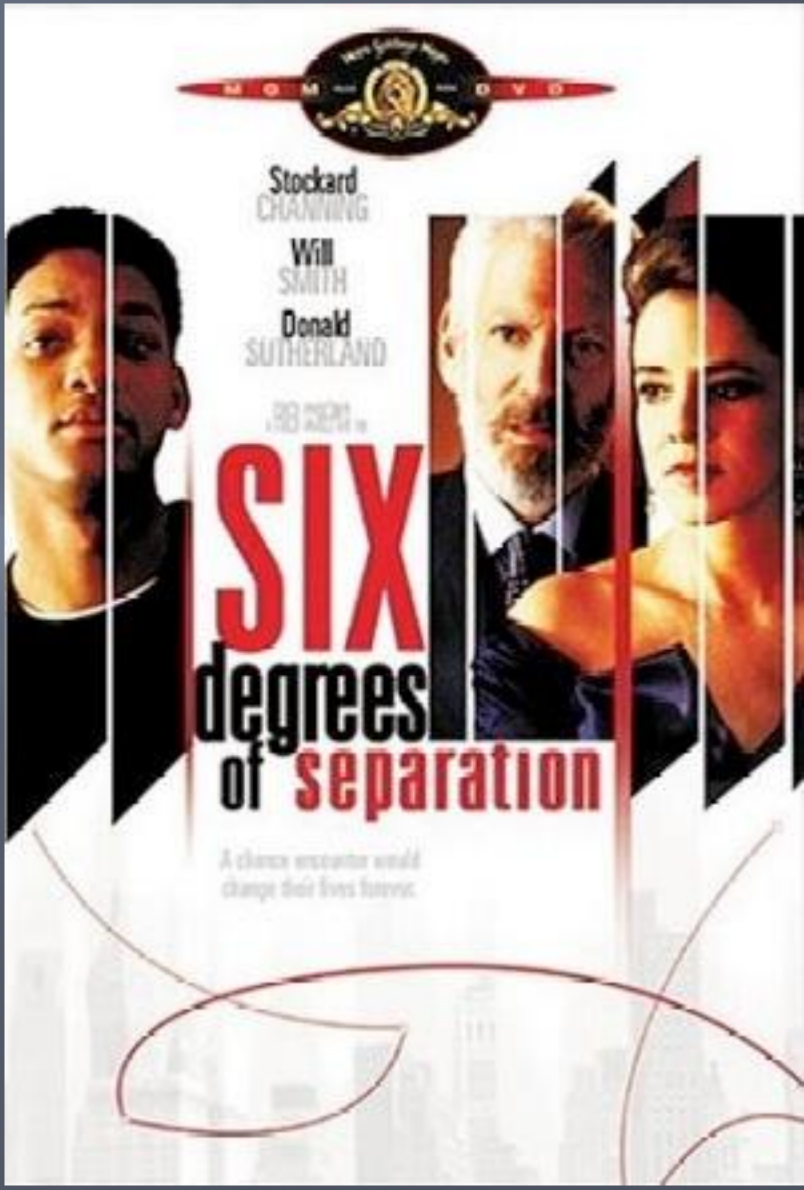
Small world

- Only small fraction reached the target person (24/120)
- Broad distributions of chain lengths (2-10), with an average of 5.5 ”Six degrees of separation”
- „Mr. Jacobs”, a clothing merkant played a key role in forwarding letters to the target
- The general scheme is that first geographic aspects dominate the search until the inner circle of the target is reached.

„What a small world!”

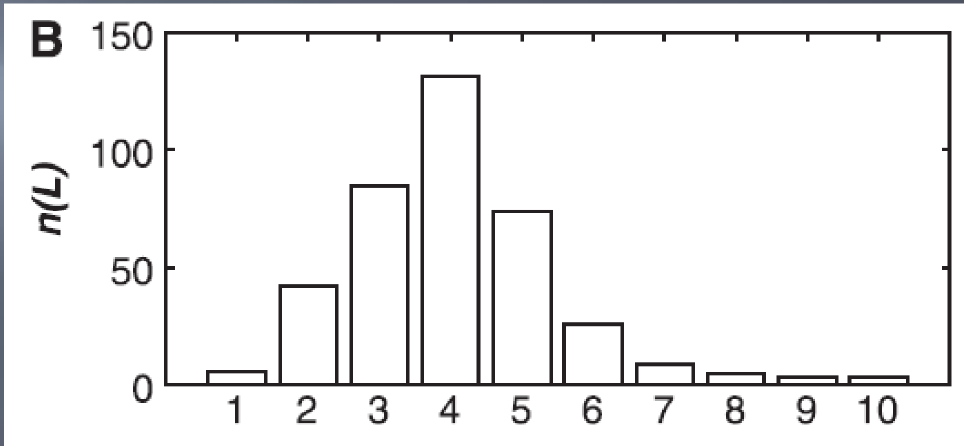
Small world

Enormous impact (see also: John Guare's play, Fred Schepisi's film, Kevin Bacon game, Erdős number)
Sixdegrees.com (1997-2001), Facebook app, LinkedIn etc



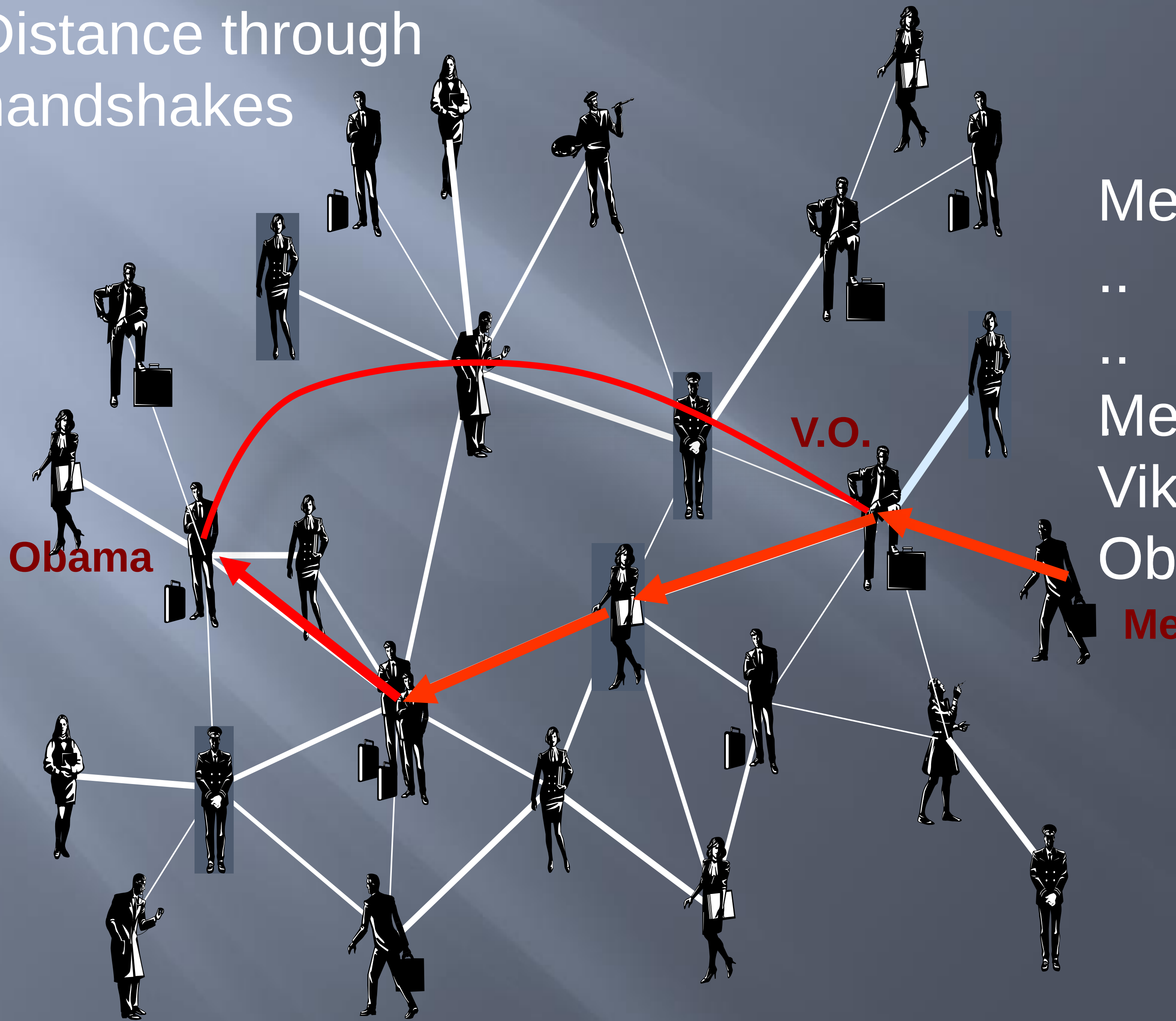
D. Watts repeated the experiment on the Internet 2003. More than 60,000 people from 166 different countries were approached in the experiment. Participants were assigned one of 18 target people.

Task: contact a specific one by sending email to people they already knew and considered potentially "closer" to the target. The targets were chosen at random, e.g., an Australian policeman and a veterinarian from Norway. 384 of 24,163 chains were completed with a mean of 4.01, median 5-6.



Small world

Distance through handshakes



Me
..
..
Me
Viktor Orbán
Obama

Small world

Mach more interesting: Kevin Bacon game

Try to find the shortest chain of actors between an arbitrarily chosen actor or actress and Kevin Bacon such, that a link between two actors is a movie where the they played together.

This is path on a bipartite graph and you are looking for the shortest one

Introduce the Bacon number B as the distance on the projected graph of actors.

B=0 Kevin Bacon

B=1 Julia Roberts, Kevin Costner, Tom Hanks

B=2, Anouk Aimée, Leonardo DiCaprio,

B=3 Judit Pogány, Charles Chaplin (!)

Small world

How do I know?

From Oracle: http://oracleofbacon.org/

Uses IMDb: Internet Movie Database
http://www.imdb.com/

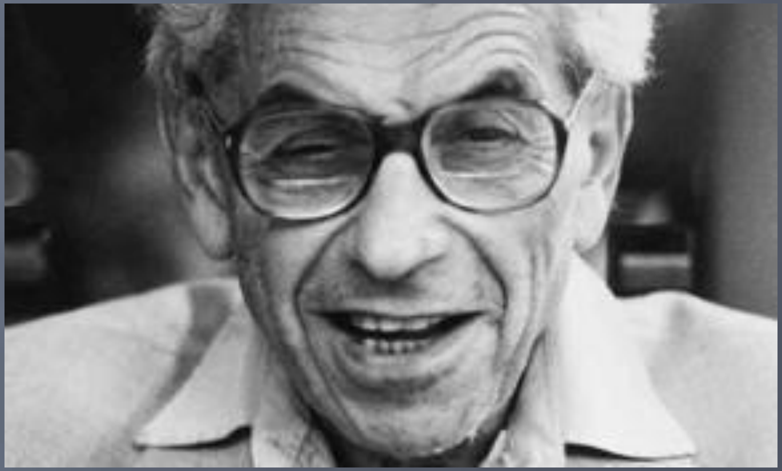
I could not find B>3, although Bacon is not that famous (except of the game!) .

But in one step you are at a famous actor.

Small world

Let us be more scientific!

An eccentric mathematician, no home, all property one suitcase, travelling around the globe and collaborating with 511 people in almost all fields of mathematics resulting in more than 1500 papers .



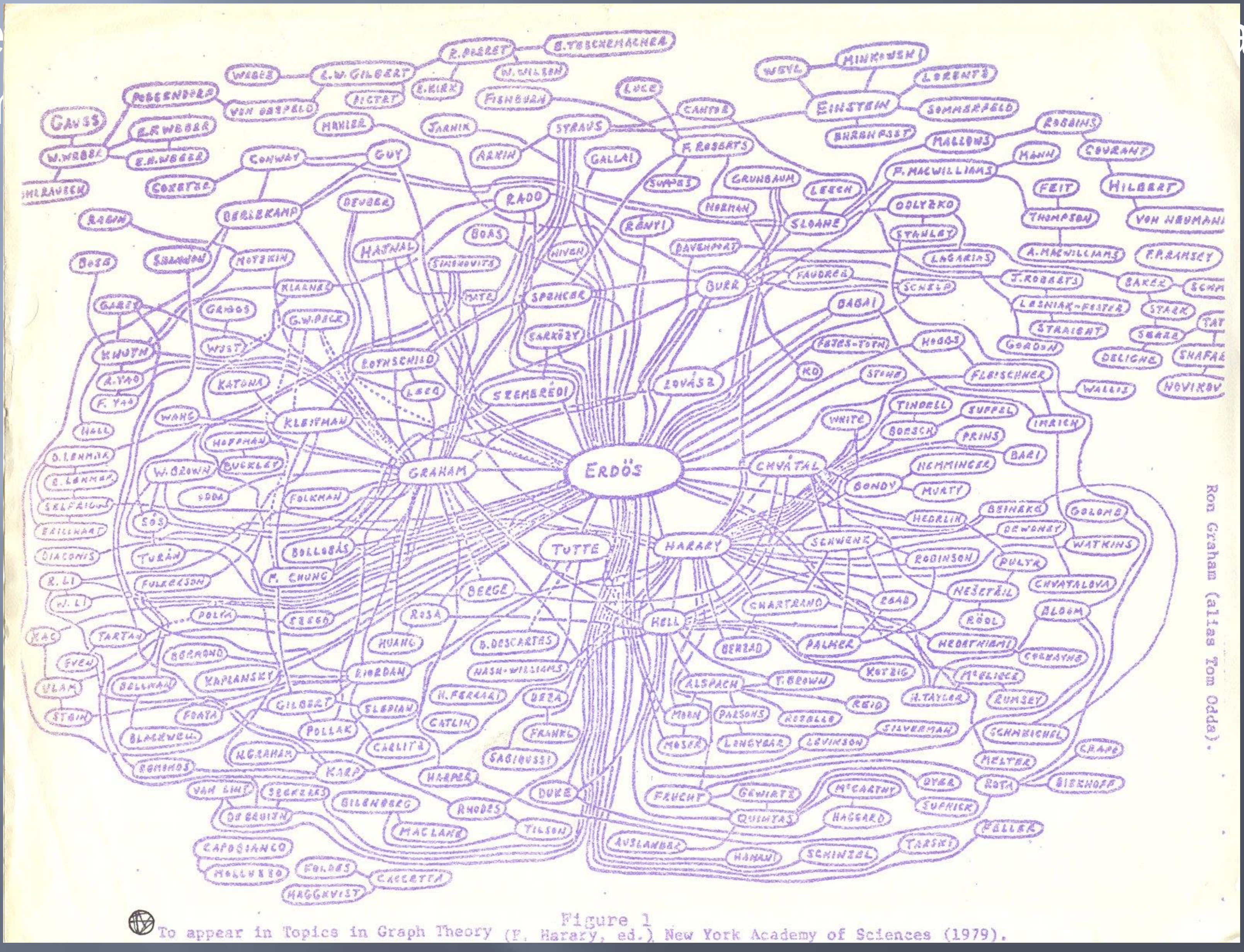
Paul Erdős (1913-1996)

Collaborative distance from Erdős on the bipartite graph of mathematicians and papers defines the Erdős number.

Small world

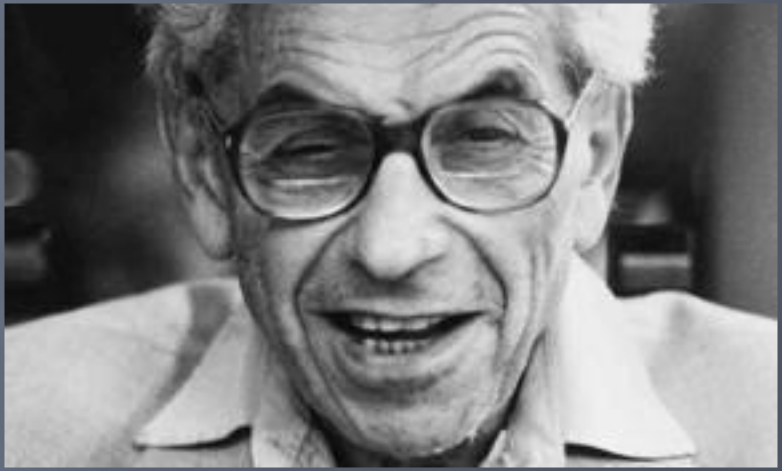
The ans

from



Small world

The Erdős graph is the projection of the collaboration graph onto the set of mathematicians (who have finite E).



Paul Erdős (1913-1996)

Most mathematicians do have a small E

Average is 4. (There are close to 10,000 with E=2)

My E?

Microsoft collaboration path checker:
http://academic.research.microsoft.com/VisualExplorer

Small world

We see on these examples that social networks are small worlds.

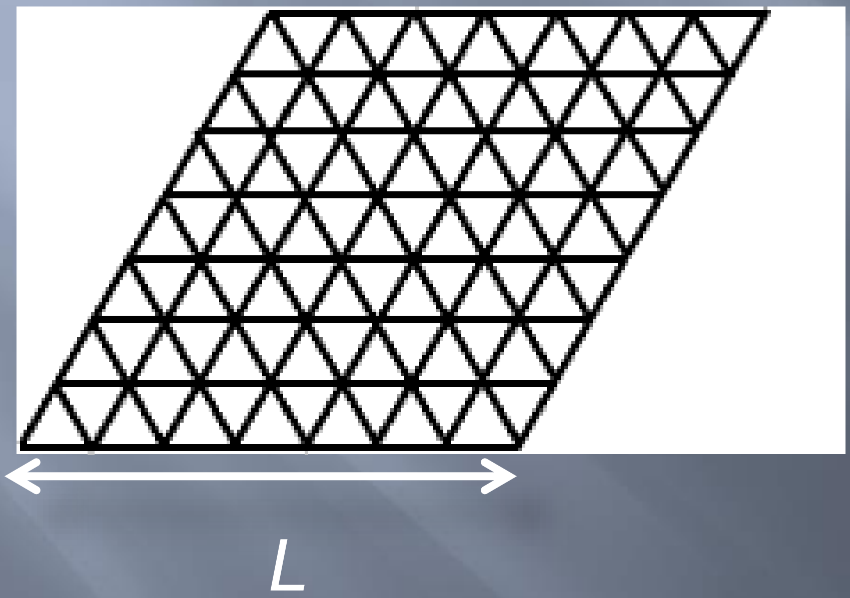
Surprisingly, many other networks in nature and technology have also small average distance.
Examples include:
- Internet, WWW
- Biochemical networks (genetic transcription, metabolic)
- Air traffic network
etc.

Universality: Is there some common mechanism?

Small world

ER graphs have small clustering!
+ narrow degree distribution

Regular lattices have absolutely narrow:



k_i = 6 for all nodes (except the boundaries)

Graph distances between distant points in 2d Euclidan sense go as $L \sim N^{1/2}$

In d dimensions $\langle d \rangle \sim N^{1/d} \gg \log N$

Not small world

Small world

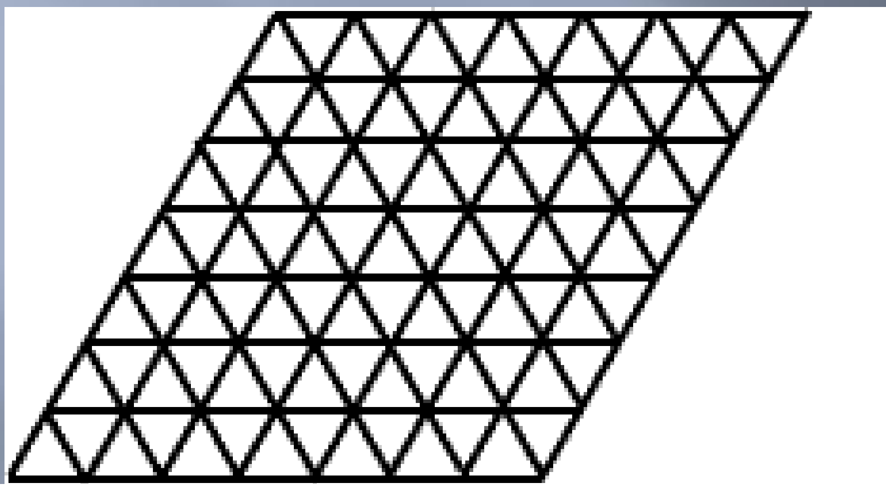
We know that the ER model leads to the small world result with (basically) constant node degrees such that it forms an exponential network without loops! (In fact, there are loops in the giant component, but close to $\langle k \rangle=1$ they are negligible.)

But there are a lot of loops in a social network! Friends of friends get often friends, if A writes a paper with both B and C then there is high chance that B will write a paper with C too etc. Triangles are important!

ER has small clustering coefficient, while real (social) netrworks have high!

Small world

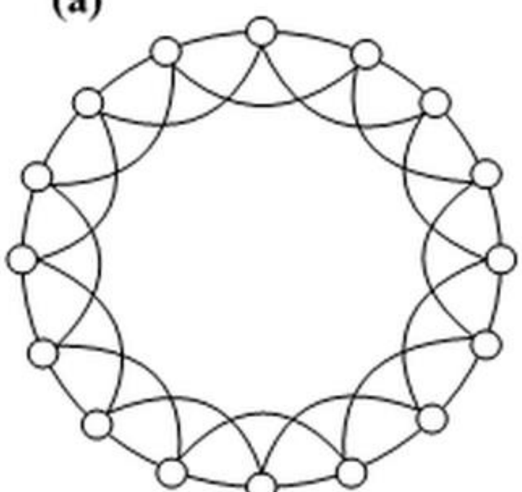
Challenge: How to match these two properties: High clustering AND small average distance.



(a)

Lattices can have high clustering. E.g. here

$$C_i = \frac{n_i^{\Delta}}{k_i(k_i - 1)/2} = \frac{6}{3\times 5} = \frac{2}{5}$$



In *d* chain with second neighbor links and periodic boundary conditions

$$C_i = \frac{n_i^{D}}{k_i(k_i - 1)/2} = \frac{3}{2\times 3} = \frac{1}{2}$$

Small world

ER:
Small clustering (bad) small average distance (good)

Regular lattices:
Large clustering (good) large average distance (bad)

We do not know how to cure ER

Can we do something with the lattices?

Watts-Strogatz model

Watts-Strogatz model (1998):

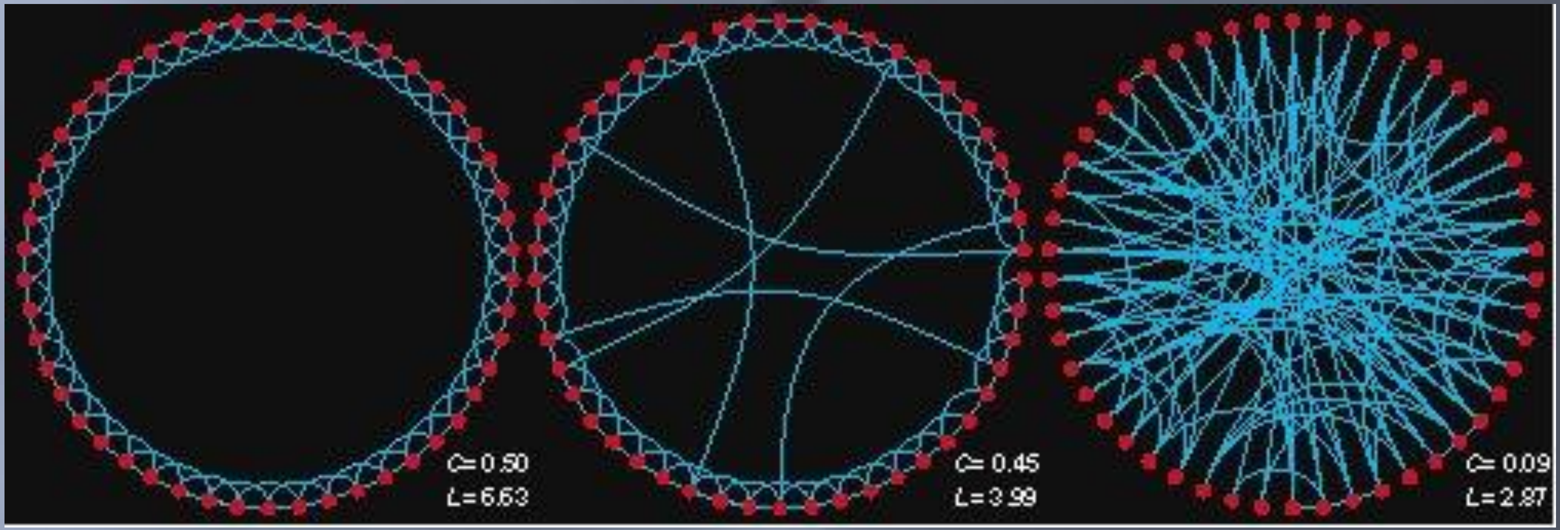
There are communities (e.g., villages), where people know each other well → high clustering

Duncan Watts

Steven Strogatz

But in the villages there are a few persons who travel, know people from other villages etc. They decrease the length of the paths.

Watts-Strogatz model



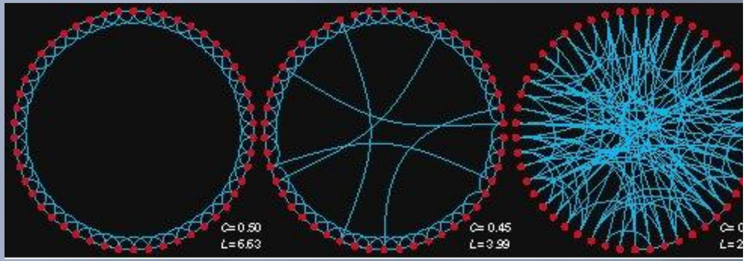
v1: WS model: Take a lattice with high clustering and rewire *p* fraction of the links. *p* → 1: ER graph

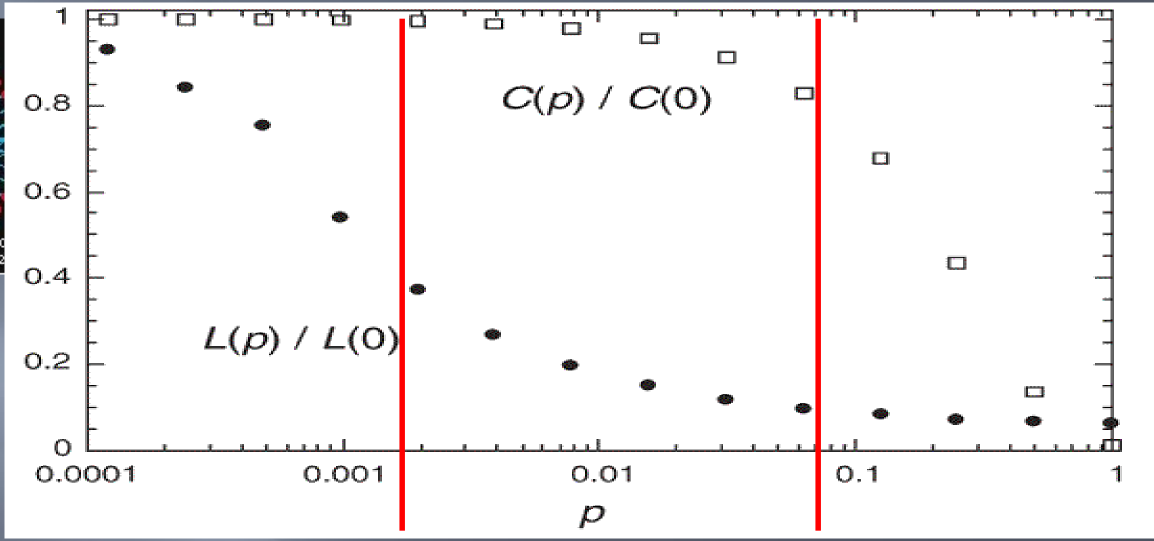
v2: Alternative version: Take the lattice and connect nodes with *p* probability. For *p* → 1 complete graph

v3: : Take the lattice, go through the links and draw a bridge to a random node with prob. *p*. Differs from v2 in large *p* limit.

Any high clustering lattice would do in any dimension.

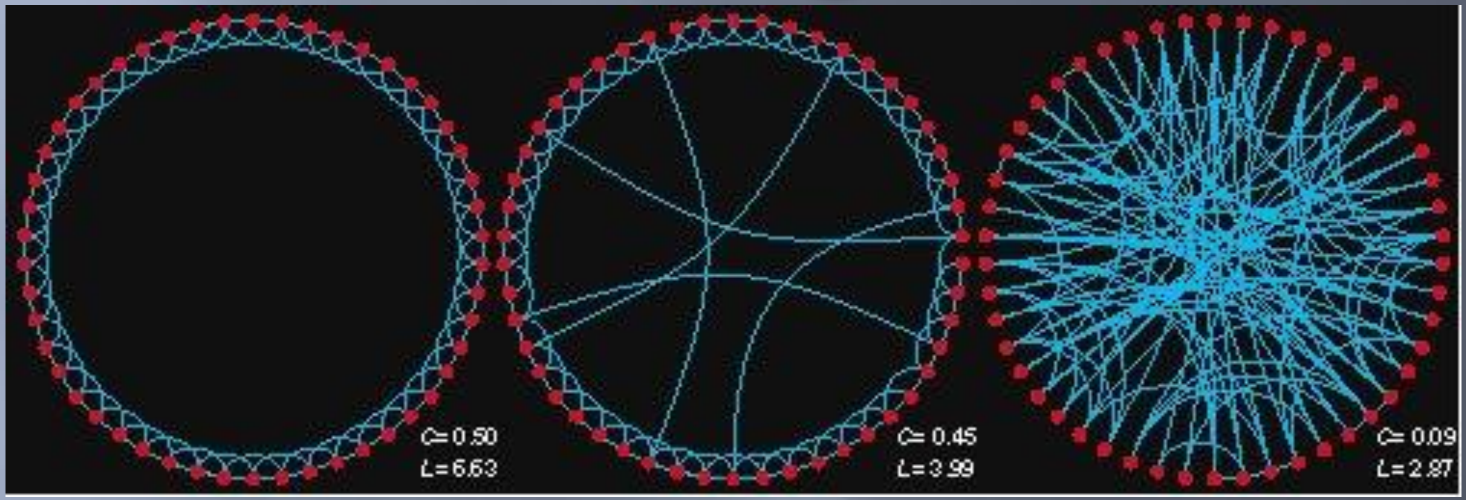
Watts-Strogatz model





Between the two red lines clustering is high and av. distance low! Problem solved! (?)

Watts-Strogatz model



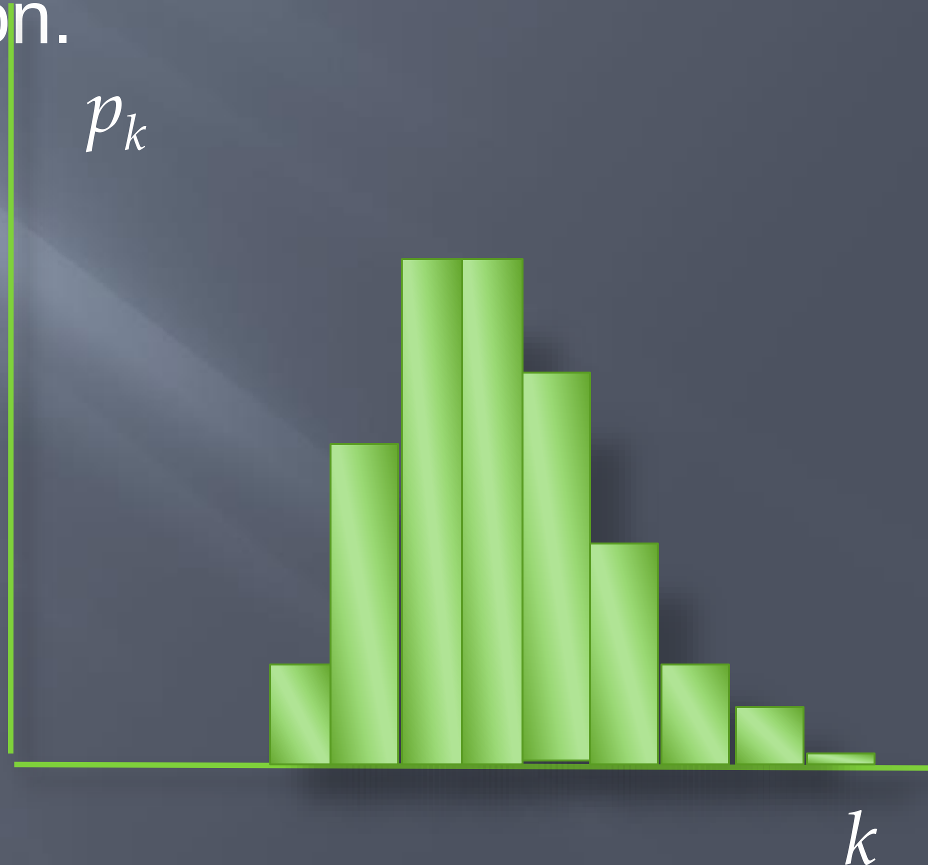
In the small world of the Watts Strogatz model people are still rather uniform. They have more or less the same number of acquaintances – only few of them have the chance to break out of this boring world.

What is the degree distribution of the nodes in the WS model?

Watts-Strogatz model

Degree distribution: Added links form an ER NW with prob *p.* If the original lattice has coordination number k_0 we finally get for the distribiution of the total degree *k* a shifted Poisson distribution.

$$p_k = e^{-<k-k_0>} \frac{<k-k_0>^{k-k_0}}{(k-k_0)!}$$



Sharply peaked, shifted Poisson

Watts-Strogatz model

Global clustering coefficient: $C = \frac{\#\text{triangles} \times 3}{\#\text{connected triples}}$

The number of Δ-s on a ring with $c = \langle k \rangle$ and *N* nodes is $N\frac{c}{4}\left(\frac{c}{2} - 1\right)$. Why? *c* is always even. It is enough to count the Δ-s on one side of the nodes, $z = c/2$.

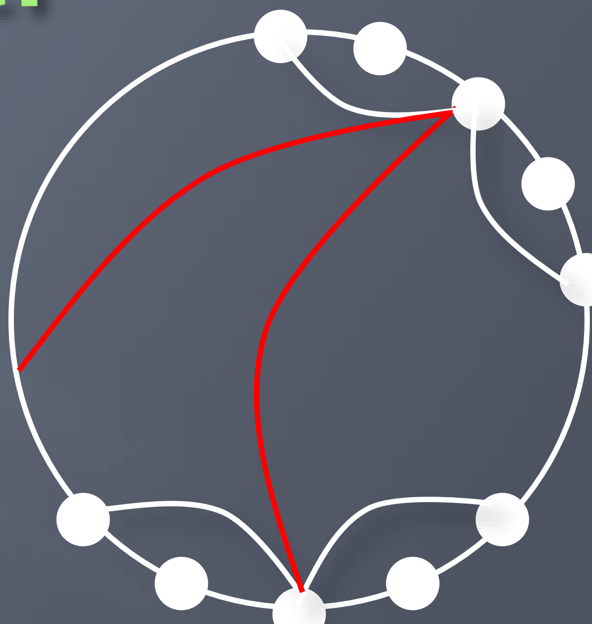
With z-th new link $z - 1$ new Δ-s. # Δ= $\sum_{I=1}^{z}(i-1) = \frac{1}{2}z(z-1) = \frac{c}{4}\left(\frac{c}{2} - 1\right)$.

Take v3 with adding bridges to the *c*-ring: For every edge in the ring we add a bridge with prob. *p.*

The expected # bridges $s = Ncp/2$, each can be put to $N(N-1)/2$ places → prob. that a pair of nodes is linked $\frac{Ncp/2}{N(N-1)/2} = \frac{cp}{N-1} \sim \frac{cp}{N}$. The probability of creating a triangle by a bridge is negligible for $N \to \infty$.

Watts-Strogatz model

$c = 4$ ring



There are $Nc(c-1)/2$ triples in the original ring, which all survive.

One bridge creates $2c$ new triples,
→bridges contribute this way by
$2c \times Ncp/2 = Nc^2p$

Another contribution comes from bridges ending at the same node.

If there are *m* such nodes at a vertex, $m(m-1)/2$ new triples are created. *m* has Poisson distribution with mean *cp* (ER):

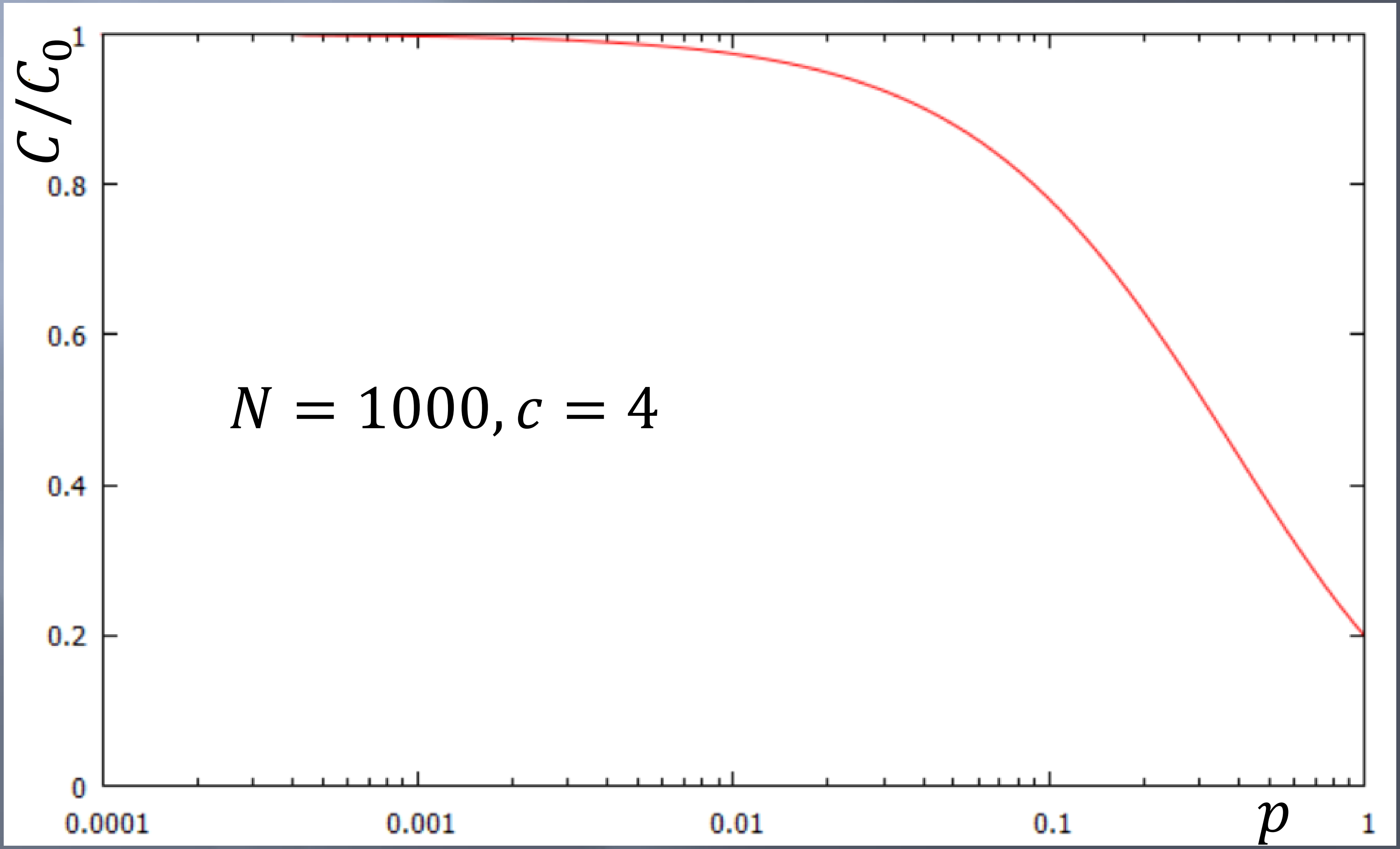
$$\frac{1}{2}[\mathbb{E}(m^2) - \mathbb{E}(m)] =$$

$$\frac{1}{2}\Big[\underbrace{\mathbb{E}(m^2) - \big(\mathbb{E}(m)\big)^2}_{cp} - \underbrace{\mathbb{E}(m)}_{cp} + \underbrace{\big(\mathbb{E}(m)\big)^2}_{(cp)^2}\Big] = \frac{1}{2}(cp)^2$$

→ $\frac{N}{2}(cp)^2$ triples

Watts-Strogatz model

$$C = \frac{\#\text{triangles} \times 3}{\#\text{connected triples}} = \frac{N\frac{c}{4}\left(\frac{c}{2}-1\right)\times 3}{\frac{Nc(c-1)}{2}+Nc^2p+\frac{N}{2}(cp)^2}$$

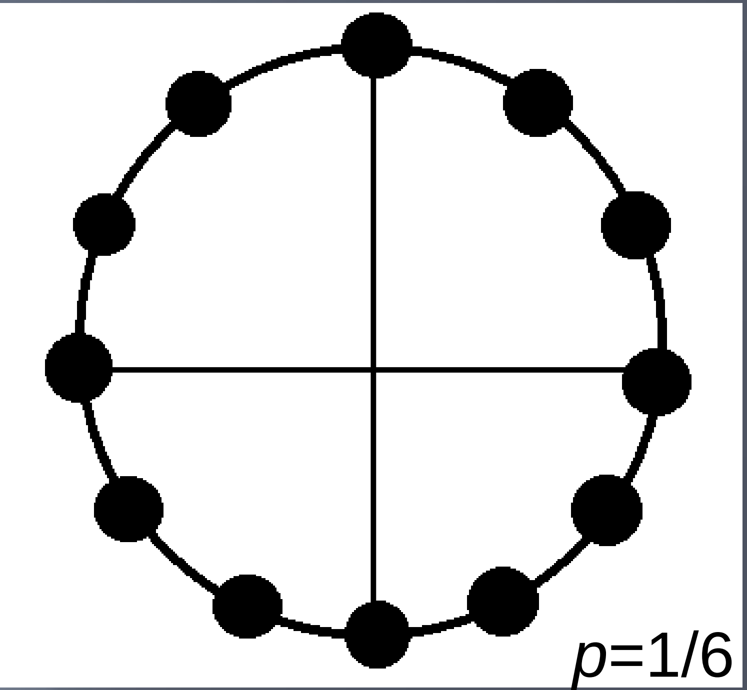


Watts-Strogatz model

Handwaving argument for diameter

First we ignore the further neighbor links - unimportant for order of magnitude estimation (and would anyway shorten δ_R).

$\delta_R \rightarrow$ we can approximate it by the length of larges gap between two neighboring bridge heads.

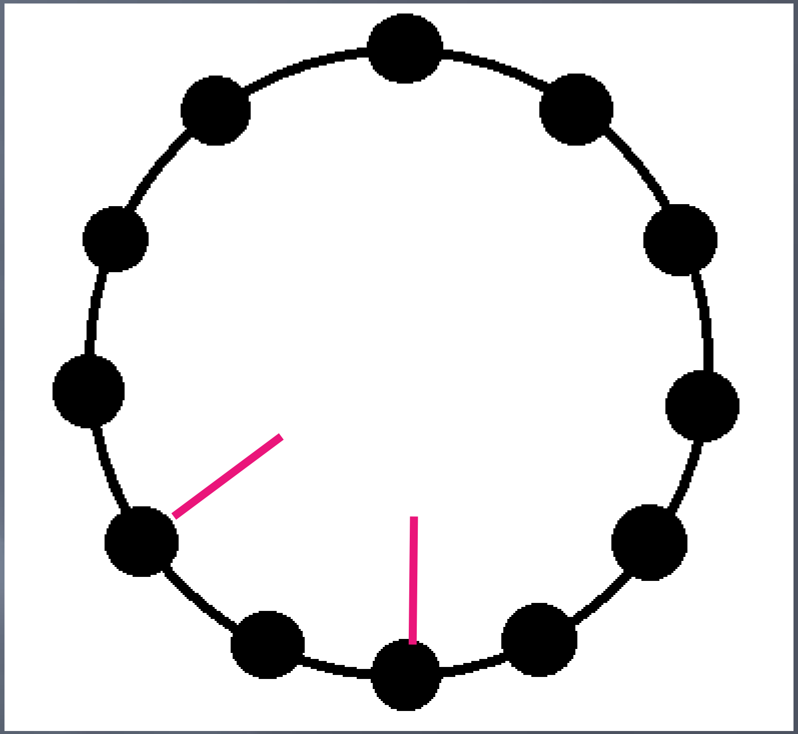


Without bridges $\delta_R \sim N$. If we put the *s* bridges regularly, i.e. with an angle of $2\pi/pN$, the largest distance would not depend on *N* ! ($\delta_R \sim N/(s/2) \sim 1/cp$.This is an underestimation as due to randomness there will be larger gaps.

Watts-Strogatz model

How to take randomness into account?

A further simplification: In the model we go around the ring and at every node we generate a bridge with prob. *p*, which, has another end possibly shortening a gap. We ignore that (and overestimate the gaps).



What is the expectation of the largest gap?
$\delta_R(N) \sim \mathbb{E}[\max(\Delta(N))]$.

First: $\mathbb{P}(\Delta) = C(p)(1-p)^\Delta$ with $C(p) = 1/\sum_{\Delta=0}^{N}(1-p)^\Delta$

Watts-Strogatz model

Large *N*: What is the expectation of the largest gap from *N* trials on an infinite ring?

$\mathbb{P}(\Delta) = p(1-p)^{\Delta}$ continuum limit →$\mathbb{P}(\Delta) = \frac{1}{w} e^{-\Delta/w}$

with $w = (1-p)/p = \mathbb{E}(\Delta)$. → $\mathbb{P}(\Delta < n) = 1 - e^{-n/w}$

Let $X_N = \max \Delta(N)$

$\mathbb{P}(X_N < x) = \mathbb{P}(\Delta < x)^N = \left(1 - e^{-x/w}\right)^N$

$\mathbb{E}[X_N]$ increases with *N*. How? If we find a function $f(N)$ so that the distribution $\mathbb{P}[X_N - f(N)]$ converges to a limit distribution: $\mathbb{E}[X_N] \sim f(N)$. Here: $f(N) = \ln N$

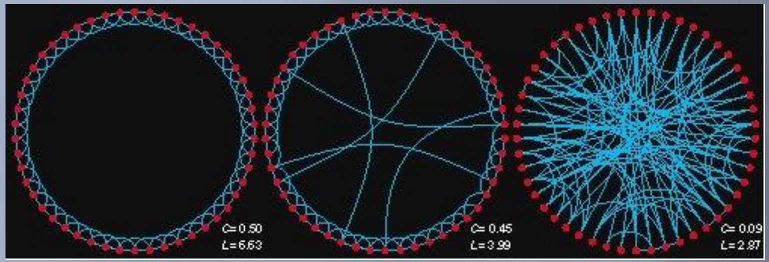
$$\mathbb{P}[X_N - \ln(N) < x] = \mathbb{P}[X_N < x + \ln(N)]$$

$$= \left(1 - e^{-(x/w + \ln(N))}\right)^N = \left(1 - \frac{e^{-x/w}}{N}\right)^N \to e^{-e^{-x/w}}$$

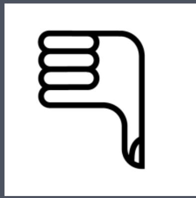
Gumbel-d

$\boldsymbol{\delta_R(N) \sim \ln(N)}$

Small world



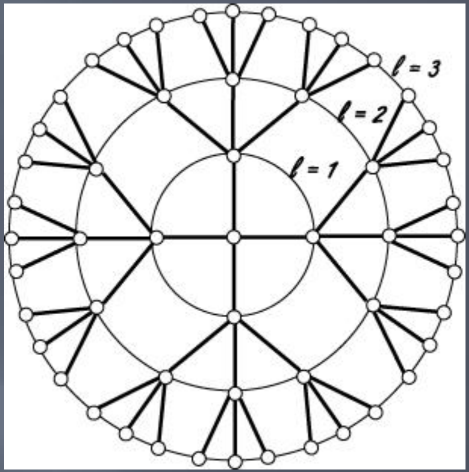
Summary of the WS model:

- Combines large clustering of some lattices with short average distance due to cross links 
- Reflects some aspects of social networks (communities with high clustering connected by long distance links). 
- It has a sharp degree distribution – in contrast with real world networks 

Different routes to small worldness

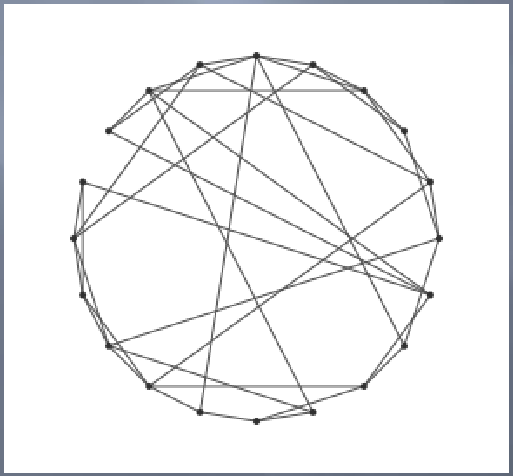
a) Exponential network, e.g., Cayley tree:

$N_t \sim \langle k-1 \rangle^t$; while $\langle d \rangle \sim t$



As ER is tree like and has sharp $P(k)$, it is in this category

b) Watts Strogatz: Bridges

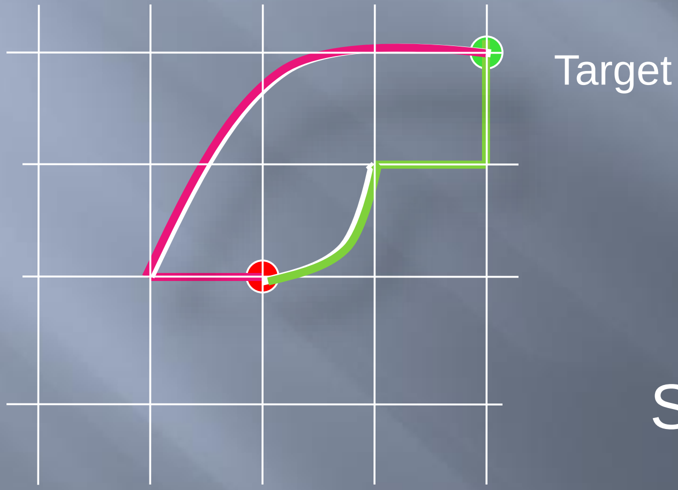


Other route?

Navigation in a small world

Navigation

Milgram’s problem is „solved” by a greedy algorithm



Greedy algorithm: 3 steps

Decentralized search: 2 steps

Shorter than 6 degrees of separation!

How many steps needed to find a target? Depends on the architecture. Average delivery time $\overline{\tau}$ as calculated from greedy algorithm for pairs of nodes.

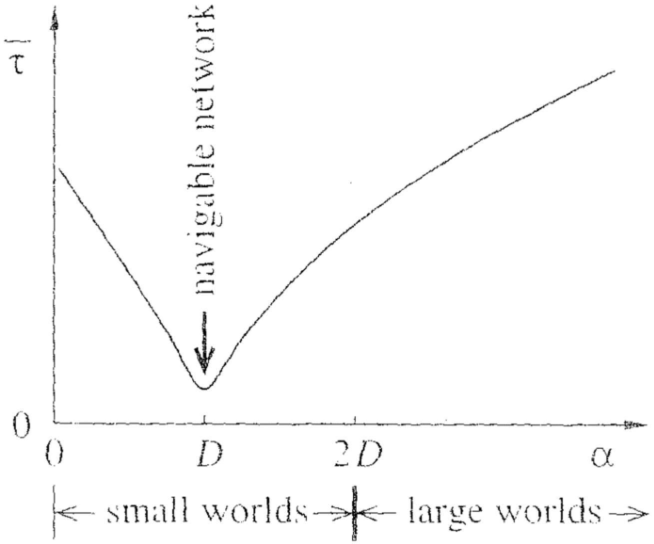
Navigation in a small world

Lattice without shortcuts: $\bar{\tau} \sim L$

Jon Kleinberg's navigability problem: How do shortcuts influence the delivery time?

1 shorcut of length ℓ is introduced with prob. $P(\ell) \sim \ell^{-a}$ for each lattice site.

There is an optimum for alpha! $\alpha = d$



$$\bar{\tau}(L) \sim \begin{cases} L^{(d-\alpha)/(d+1-\alpha)} & 0 \le \alpha < d \\ \ln^2 L & \alpha = d \\ L^{\alpha-d} & d < \alpha < d+1 \\ L & \alpha > d+1 \end{cases}$$

Only at navigability point is greedy algorithm good!

Homework:

Generate WS graphs from rings with first and second neighbor connections (version 2: add links between nodes with probability p) with $p = 0.05$ and different N-s.
Calculate average distance d_D of diametric points.

For every value of N many samples are needed.
Go with N to possibly large sizes.

How does d_D depend on N?